VOL. XVIII FEBRUARY, 1934 No. 1

The Journal of APPLIED PSYCHOLOGY

JAMES P. PORTER, EDITOR
Ohio University
WITH A BOARD OF CO-OPERATING EDITORS

TABLE OF CONTENTS

HENRY C. LINK. A New Method for Testing Advertising and a Psychological Sales Barometer 1

KATHARINE VAN TUYL and ALVIN C. EURICH. Measuring the Interests of College Students with Different Major Subjects 27

FRANK N. STANTON. Memory for Advertising Copy Presented Visually *Versus* Orally 45

E. B. SKAGGS. A Summary and Critique of Methods of Measuring "Warming-Up" 65

PAUL J. FAY and WARREN C. MIDDLETON. An Economical Method of Administering and Scoring New-Type Examinations 77

GEORGE J. DUDYCHA. The Beliefs of College Students Concerning Evolution 85

PAUL R. FARNSWORTH. Familiarity with the Names of Musicians 97

JOHN W. DICKEY. On Estimating the Reliability Coefficient 103

SELMA M. MATHEWS. The Effect of Mothers' Out-of-Home Employment upon Children's Ideas and Attitudes 116

B. F. HAUGHT. Mental Growth of the Southwestern Indian 137

W. N. BRIGANCE. The Language Learning of a Child 143

News and Notes 155

Book Reviews 157

New Books and Pamphlets 163

PUBLISHED BI-MONTHLY BY THE EDITOR
Ohio University, Athens, Ohio
Printed by
C. E. PAULEY & COMPANY, INC.
225 N. New Jersey St., Indianapolis, Indiana

Entered as second-class matter at the post office at Athens, Ohio, under the act of March 3, 1879. Additional entry as second-class matter at the post office, Indianapolis, Indiana.

COPYRIGHT 1933, BY JAMES P. PORTER

The Journal of Applied Psychology

CO-OPERATING EDITORS

W. F. BOOK, Indiana University, W. V. BINGHAM, Personnel Research Federation, New York; M. E. BROOM, State Teachers College, San Diego, Calif.; ROBERT H. GAULT, Northwestern University; WILLIAM HEALY, Baker Foundation, Juvenile Court, Boston; H. L. HOLLINGWORTH, Columbia University; J. E. W. WALLIN, Atlantic University; HERBERT S. LANGFELD, Princeton University; JAMES H. LEUBA, Bryn Mawr College; HENRY M. MCCOMAS, Princeton University; WALTER D. SCOTT, Northwestern University; C. E. SEASHORE, University of Iowa; DANIEL STARCH, Boston, Mass.; EDWARD K. STRONG, JR., Leland Stanford Jr. University; L. M. TERMAN, Leland Stanford Jr. University; E. L. THORNDIKE, Columbia University; JOSEPH JASTROW, New School of Social Research, New York; G. M. WHIPPLE, Danvers, Mass.; HELEN THOMPSON WOOLLEY, Teachers College, Columbia University.

The scope of THE JOURNAL OF APPLIED PSYCHOLOGY will include the following: (a) The application of psychology to vocational activities, such as law, art, public speaking, industrial and commercial work, and problems of business appeals. (b) Studies of individual mentalities, such as types of character, special talents, genius, and individual differences, including the problems of mental diagnosis and vocational prognosis. (c) The influence of general environment conditions, such as climate, weather, humidity, temperature; also such conditions as nutrition, fatigue, etc. (d) The psychology of everyday activities, such as reading, writing, speaking, singing, playing games or playing musical instruments, sports, etc.

Communications, manuscripts, reviews, summaries, notices, books and off-prints for review should be addressed to Dr. James P. Porter Ohio University, Athens, Ohio. Contributors are requested to follow the conventional method of citation employed by this JOURNAL; in every case the year, volume and pages should be specified.

THE JOURNAL OF APPLIED PSYCHOLOGY is published bi-monthly. The subscription price is $5.00 a year. Single numbers $1.00 to $2.00, depending upon the number of pages. Remittances and business communications should be addressed to the Editor, James P. Porter, Ohio University, Athens, Ohio.

A NEW METHOD FOR TESTING ADVERTISING AND A PSYCHOLOGICAL SALES BAROMETER

Developed Through Experiments Among 14,000 Consumers, by 60 Representatives of the Psychological Corporation

Described by

HENRY C. LINK, Ph. D.

New York, N. Y.
December 21, 1933.

Dear Dr. Porter:

In view of the unique character of these studies, involving as they do the voluntary cooperation of some 60 psychologists and several hundred of their students, a word of explanation as to their origin and conduct is in order. As managing director of the Psychological Corporation, this gives me a peculiar pleasure, because these studies demonstrate that the vision of the eminent psychologists who founded the Psychological Corporation in 1921 was truly prophetic. Aside from the specific merits of the studies themselves, they illustrate the achievement of an ideal expressed in the Corporation's charter, namely, that psychologists would cooperate in scientific studies of a practical nature even to the extent of contributing their time and effort without immediate financial remuneration. Indirectly, these studies have led to other studies paid for by business organizations; but taken by themselves, they represent the expenditure, in commercial equivalents, of about $25,000. The results of this research, through this and other articles, is made available to all those interested in scientific market research in its psychological aspects.

In the fall of 1931, my associate, Dr. Henry C. Link, began a series of small experiments to try out the possibilities of different advertising tests. The distinctive feature of these tests was that they were made on homogeneous groups of consumers, and dealt only with matters in which these consumers had a practical interest. Housewives, for example, were tested on advertising having to do with household products. Men were tested on advertising having to do with suits, tobaccos, shaving materials, and products which they might be expected to buy.

The decision to confine these studies to actual or potential buyers of the articles advertised grew out of Dr. Link's analysis of advertising studies made in the past (See Link's recent book on *The New Psychology of Selling and Advertising*, Macmillan, 1933) from which he concluded that tests of advertising could give reliable results only if made on the people for whom the advertising was intended. This meant tests in the home rather than laboratory tests.

The half dozen preliminary experiments were so promising that I encouraged Dr. Link to invite the cooperation of our representatives in trying the best of these tests on a larger scale. The test chosen for these studies is one which Dr. Link designates as the method of *triple associates*. The first of these larger studies was described in the Harvard Business Review (January 1933) and fragmentary results of subsequent studies have appeared in *Printer's Ink*, *Advertising and Selling*, and various business publications. The interest aroused by these articles was extremely helpful in lending direction to the studies being made, besides serving a useful purpose in a field currently beset by all kinds of commercial short cuts and unsound psychological theories and practices.

This article is the first complete presentation of the method of triple associates and the concomitant result of these studies, a psychological sales barometer.

Incidentally, letters from about thirty psychologists tell us that these studies have added a wealth of new material and new activity to their courses in applied psychology. We have not been able to report the results of all studies in detail to the psychologists who participated, because of the heavy expense involved. In this statement, however, I wish to express the official appreciation of the Directors and officers of the Psychological Corporation for the cooperation of the many psychologists who have participated in these studies. This appreciation has been concretely expressed in the election to the Board of Directors last month, of Dr. J. B. Miner, head of the Department of Psychology, University of Kentucky, and Dr. Rensis Likert, New York University, as assistant secretary and treasurer. Dr. Miner and Dr. Likert have been among our most active market representatives.

PAUL S. ACHILLES
Managing Director
The Psychological Corporation
New York, N. Y.

A. The Method of Triple Associates

When an advertiser spends thousands and even millions of dollars in repeating a theme like "Look for the Date on the Can" (Chase and Sanborn Coffee), or "Fire-chief Gasoline" (Texaco), or "Floating Power" (Chrysler automobiles) it is not for the sake of impressing that theme on buyers' minds. The advertiser's purpose is to impress his *product*. He chooses such a theme as a psychological messenger or *vehicle* with which to carry the name and identity of his product into the minds of possible purchasers. This vehicle, or *advertising idea* as it is called, is important only insofar as it registers or fixes the product. When Ipana toothpaste advertises "Pink Toothbrush" it is not trying to sell pink toothbrushes but Ipana toothpaste. When Texaco gasoline advertises "Fire-chief" the primary purpose is to register not *"Fire-chief"* but *Texaco gasoline.*

The use of a theme or association vehicle, such as those mentioned, represents what psychologists have long agreed on as the essential mechanism of successful advertising. This mechanism consists of a deliberate attempt to establish an association between some interest, want, habit, or emotion which the buyer already has, and the name of the particular product advertised. "If an advertiser," says Poffenberger (*Psychology in Advertising*, 1925, p. 483) "demands that the name of his particular soap shall come to mind when the need for such an article is felt, it would seem that the only necessary condition to be fulfilled is that these two ideas—namely, the need for soap and the trade name of the particular brand of soap—shall have been associated in the mind previously.—The advertiser is, therefore, faced with two questions, namely:

"1. How can two ideas be associated in the mind?

"2. How may he guarantee that his idea (trade name) shall be the one out of all possible trade names

that shall come to mind?" Insofar as the conclusions of psychologists regarding the fundamental mechanism of successful advertising are valid, the advertising themes mentioned above are exemplifications of the application of these conclusions. The ethical values involved are a subject of major importance, but statements by psychologists, as psychologists, on this subject must naturally follow rather than precede the establishing of reliable measuring techniques.

The test to be described here consists of nothing more than putting the *process* of advertising into the form of a test question. For example, What toothpaste advertises: "Good for Pink Toothbrush?" When 62 per cent of the women tested with this question answered *Ipana,* we know that this advertising theme has succeeded in establishing the *intended* association in the minds of that many women. When 21 per cent of the women tested with the question: What toothpaste advertises "Removes all 7 stains"? answer *Colgate's,* we know that this theme has been successful to that extent.

Whether an advertising theme or vehicle has been used once in one advertisement, or a hundred times in varying forms in many advertisements, the test is the same in form. It consists of naming two members of an association process and asking the consumer to name the third. For example, *What coffee*—the first associate—*advertises: "Look for the Date on the Can"?*—the second associate. The third associate is the answer given, *Chase & Sanborn* being the correct answer in this case. Or, *What drug product* (1) advertises: *"What happens in these water glasses happens in your stomach too"?* (2). Answer: *Bayer's Aspirin* (correct) or the other names given, the third associate.

Paired Associates.

The method of *triple associates* is a logical development from the method of paired associates which psy-

chologists have found, previously, the most significant of all verbal tests in studying advertising. The method of *aided recall* or paired associates, according to Poffenberger, "is the one that most nearly measures memory for advertisements and trade names as it actually occurs in every day life." In his comprehensive analysis of advertising tests, particularly the recognition tests, aided recall tests, and unaided recall tests, Dr. Poffenberger states: "It is obvious from the foregoing discussion that the choice of the method that shall be used for measuring the memory of advertisements is not entirely a matter of convenience. Since the results obtained from different methods vary as they do, it will be necessary to inquire how one remembers or rather recalls an advertisement and how such memory can be most adequately measured. When one recalls an advertisement or the trade name of a product which has been advertised there is a reason for his doing so. There is something in his mind with which that advertisement or trade name has been associated. Indeed, it is the function of the advertising to establish just such associations, the more of them the better. But the most important association or bond that can be established is that between a desire or need and some specific means of satisfying the desire or need. Thus, when one feels a need or desire to own a vacuum cleaner, that brand which comes to mind and succeeds in staying there will stand a good chance of being purchased. Now this form of memory is neither pure recall, such as is involved in reciting a poem, nor is it recognition, such as is involved in identifying an advertisement as seen before. It resembles most closely what has been called 'aided recall' in this chapter, or the laboratory method of 'paired associates.' The first member of the pair is the need or desire, and the second member is the particular product that has been shown by advertising to satisfy the need or desire." (*Ibid,* p. 519.)

Instead of asking, as in the method of paired associ-

ates for example, What gasoline did you see advertised? —the method of triple associates asks: What gasoline advertises "Fire-chief" gasoline? The added associate, *Fire-chief*, is the specific mechanism used to associate Texaco gasoline with the desire for a gasoline of known performance.

After several small preliminary tests had indicated the possibility of measuring wide differences in the percentage of correct answers to test questions of this kind, the first large scale test was tried in March, 1932, among 1578 housewives in fifteen cities. The results from 27 questions included in this test fell between the following extremes: (A complete account of the results of this study was given in the *Harvard Business Review*, January, 1933.)

TABLE I

What coffee advertises: "Look for the DATE on the Can"?

Answers	No.	Percent
Chase & Sanborn (correct)	1090	69.1
Maxwell House	81	5.1
Miscellaneous brands	74	4.7
Don't know	333	21.1
Total	1578	100.0

TABLE II

Q. What building company advertises "I smell smoke"?

Answers	No.	Percent
Johns-Manville (correct)	60	3.8
Celotox	12	.8
Miscellaneous	88	5.6
Don't know	1418	89.8
Total	1578	100.0

The theme represented by the first test question had been used in national magazines, radio, newspaper, and

display advertising for over a year. The theme "I smell smoke" had been used in a single advertisement in the *Saturday Evening Post,* as the text of a story about asbestos shingles. Incidentally, the result of the test on this theme is one of many which provides clues as to the small extent to which guessing is a factor in these tests. Even in respect to articles widely used, as many as 85 per cent and 90 per cent of the people tested answer in some tests: "I don't know."

Statistical Reliability.

The first large scale study demonstrated the reliability of this method so far as the consistency of results obtained in different localities was concerned. A glance at table III will quickly reveal this consistency. Statistically we may treat the answers to a given question as a binomial distribution, in which all the correct answers are one series and all other answers, including the *don't knows,* are the second series. On this basis the sigmas and probable errors are as given below.*

* Considering the reliability of the 15 different samples as related to the result from the total number of cases, the probable error of the Mean is: Table IIIa, 0.34 per cent; Table IIIb, 0.13 per cent; Table IIIc, 0.28 per cent. The probable errors, based on the standard deviations of the localities, are: Table IIIa, 13.36 per cent; Table IIIb, 5.19 per cent; Table IIIc, 8.84 per cent. The standard deviations on which these probable errors are based are derived as follows: $\sigma = \sqrt{\frac{\Sigma\ F\ D^2}{N}}$ in which F is the frequency of each sample, D the deviation of each sample from the mean, and N the total number of cases.

Considering the samples as a whole, namely, the total number of people regardless of locality, the probable errors are: Table IIIa, 0.78 per cent; Table IIIb, 0.52 per cent; Table 111c, 0.85 per cent. This is based on the assumption that we have a binomial distribution so that the standard deviation equals the square root of PQ, where P equals the relative frequency of correct answers and Q the relative frequency of all other answers divided by N, the total number of cases.

TABLE III

Answers, by Percentages, of 1,578 Housewives, by Locality

	Chicago, Ill.	Chicago, Ill.	Ames, Iowa	Brooklyn, N. Y.	Buffalo, N. Y.	Cleveland, Ohio	Denver, Colo.	Easton, Pa.	Lexington, Ky.	Milwaukee, Wis.	Missoula, Mont.	S. Hadley, Mass.	State College, Pa.	St. Louis, Mo.	Syracuse, N. Y.	Totals
Number of interviews	100	400	100	75	100	100	57	100	100	32	100	79	70	100	65	1,578

TABLE IIIa

What coffee advertises "Look for the Date on the Can"?																
Chase & Sanborn	67%	73%	10%	91%	99%	90%	68%	70%	61%	82%	52%	57%	87%	63%	74%	69.0%
Maxwell House	2	8	----	----	----	2	2	5	13	----	9	11	2	----	6	5.1
Miscellaneous	5	6	6	3	1	2	2	6	4	9	4	4	4	8	4	4.7
Don't know	26	13	84	6	----	6	28	19	22	9	35	28	7	29	16	21.1

TABLE IIIb

What company's product has a nutritional rating of triple plus?																
Kraft Phenix	5%	22%	2%	5%	5%	8%	2%	15%	5%	3%	7%	2%	16%	4%	9%	10.5%
Velveeta	----	16	----	6	----	4	----	----	3	----	----	1	3	----	----	4.9
Grapenuts	2	4	----	----	3	----	----	----	----	----	----	----	3	----	----	1.6
Miscellaneous	11	8	25	9	5	2	7	9	9	----	11	10	11	5	6	8.6
Don't know	80	50	73	80	87	86	91	79	83	97	82	87	67	91	85	74.5

TABLE IIIc

What soap advertises "Stop those Runs in Stockings"?																
Lux	50%	33%	36%	57%	44%	76%	44%	45%	51%	28%	74%	47%	56%	62%	51%	47.8%
Ivory	3	16	8	7	19	----	12	4	11	----	7	----	6	----	3	7.9
Palmolive	2	22	6	----	5	2	----	9	3	----	----	5	----	2	----	7.4
Miscellaneous	2	1	3	3	2	4	3	2	----	3	2	4	----	3	3	2.2
Don't know	43	28	47	33	30	18	41	40	35	69	17	44	37	43	43	34.4

The consistency of these results permits us to conclude:

1. That we have here a test and a procedure sufficiently definite and controlled to produce consistent results in different localities if the existing conditions warrant such consistent results. This procedure included the tabulation of the results locally, so that the total results, as compiled in New York, represent as many separate investigations as there are psychologists participating. Other details of the procedure will be described later.

2. That we are testing forces so universal in scope as to give us consistent results when rightly tested. No attempt was made in the first study to select the women tested by any criterion except that they be women who did their own housekeeping. Indeed, beyond this criterion, the selections were left purposely to chance so that the usefulness of this test might be measured under the severest possible strain.

The significance of this method, it was believed, would be better established by a wide variety of tests, given at different times and to different groups of women, than by the intensive analysis of just one set of results. Over a period of a little less than two years, therefore, the following studies were made:

TABLE IV

Time of Study	No. of Cities and Towns	No. of People
1. March, '32	14	1578
2. September, '32	18	541
3. October, '32	47	1445
4. December, '32	43	1958
5. February, '33	39	1399
6. April, '33	23	1348
7. May, '33	35	1234
8. October, '33	40	1932
9. November, '33	40	2386
Total		13821

In addition, some of the test questions were used as controls in commercial studies made among both men and women, so that we have results like the following:

TABLE V

Men and Women, December, 1932

	Per Cent of Correct Answers: Men	Women
What gasoline advertises: Fire-chief Gasoline?	70.3	56.2
What coffee advertises: Look for the Date on the Can?	70.6	75.3
What toothpaste advertises: Pink Toothbrush?	38.	62.
What automobile advertises: Floating Power?	83.1	55.4

The consistency of the results obtained at different times, among different groups of women, and even in groups of localities which were not the same, is illustrated below.

TABLE VI

Q. What Coffee Advertises: "Look for the Date on the Can"?

Answers	March 1932 %	Sept. 1932 %	Oct. 1932 %	Dec. 1932 %	Feb. 1933 %
Chase & Sanborn (correct)	69.1	66.8	71.	75.3	74.1
Maxwell House	5.1	4.6	4.1	4.5	5.1
Miscellaneous	4.7	4.4	4.1	3.0	3.7
Don't know	21.1	24.2	20.8	17.2	17.2
No. of housewives	1578	541	1445	1958	710
No. of cities	14	18	47	43	39

Among different groups of housewives, in different collections of cities, and with samples as small as 541 and 710, these results are nevertheless consistent. The marked increase in correct answers between September and December coincides with the introduction of a leading performer in the radio advertising program.

Tests of Ideas.

In all the above tests, the test questions were repeated in exactly the same words in succeeding studies. How-

ever, in other cases the wording of the test questions was altered to conform to certain variations introduced in the advertising theme. For example, a campaign for Lux flake soap, based on the theme that *runs in stockings* are being reduced through the use of Lux, was tested on four different occasions as follows:

TABLE VII

What Flake Soap Advertises: (Complete Question as Below)

Date of Study	Test Questions	No. of Women	Per Cent of Correct Answers
March, '32	"Stop those Runs in Stockings"	1578	47.7
Sept., '32	"It couldn't have happened at a worse moment, that awful stocking run"	541	53.6
Feb., '33	"College girls in one leg test cut down stocking runs 75%"	691	70.9
Oct., '33	"Cut down or stop those runs in stockings"	1932	67.5

After five tests on the theme: "Look for the Date on the Can," a sixth test was worded: What coffee advertises: "Before *dated* coffee I refused a second cup—but now I drink five cups a day"? This represented a variation in the basic theme of Chase & Sanborn coffee. The correct answer was given by 75.6 per cent of the women tested, comparing with 74.1 per cent correct answers given by an equal number of women at the same time to a question on the old theme. Such results demonstrate that the answers to test questions depend not so much on the exact words used to express the advertising theme as on the fidelity with which the central idea of the theme is maintained, throughout the variations introduced. In other words, we are dealing with association patterns that are *ideational* rather than purely *verbal.*

This conclusion has an important bearing on the objection sometimes raised that these tests are tests of headlines and slogans only. This is not a fact because in

many cases the headlines as given are not used in the test. Frequently they are amplified in the test so as to represent the *theme* more completely. In the cases where exact titles or slogans are used, it is usually because they represent the basic idea which the subsequent copy elaborates. In nearly all instances where the exact title has been used in the test question, the title has summed up the copy by which it was elaborated.

Details in Technique and Procedure.

Probably the most important detail in the giving of these tests is that the test questions are put orally to the person being tested and in exactly the same words as given. Again and again our instructions have emphasized the importance of *reading* the questions rather than trying to repeat them from memory. Emphasis has also been placed on the direction that under no circumstances are the test sheets to be left in the home to be called for later.

After the first study, many of the student investigators asked their supervisors how long they should wait for an answer before going on to the next question. It has been the rule since to allow a rough count of ten seconds before going on with the following question.

In one of the earliest studies the order of the questions was reversed so that one-half the women tested would get one question first, the other half would get that question last. The maximum difference found between the answers given to the same questions—asked in two different orders, was 2 per cent, and the average difference was less than 1 per cent. That is, if 26 per cent of all the women answered the first question correctly, probably 25 per cent answered the same question correctly when it was the 20th question in the series. Since the average number of questions asked was 25, this means that the position of a question in the series, whether at the beginning or the end, was not very important. This

might not be true of longer questionnaires, or even of shorter questionnaires when not properly constructed.

In some studies we asked questions about both the advertising of a product and the purchase of that product. By using two sets of questionnaires among an equal number of women, one set in which this was done and one set in which related questions were carefully avoided, we discovered that combining questions of use with tests on advertising did not alter the results in any significant fashion.

These are a few of the control devices introduced to give information in regard to the reliability of the tests. Others were introduced by design, and a few occurred by accident.

Effects of Repetition.

One of the most interesting results of these tests is illustrated in the following table, which shows the proportion of correct answers given to the same tests in different studies.

TABLE VIII

	March, '32 Per Cent	Sept., '32 Per Cent	Oct., '32 Per Cent	Dec., '32 Per Cent	May, '33 Per Cent
Test A	18.3	13.1			13.9
Test B	28.			23.1	18.2
Test C		10.2	10.		

The advertising theme represented by Test A had been running in major magazines for almost three months when first tested, and was used consistently throughout the period of succeeding tests. The theme represented by Test B was also used extensively, and with extra volume during the few months preceding the last test. Nevertheless, both themes show a decline in effectiveness. The theme in Test C was extensively used during the months of our tests, yet shows no cumulative effect due to an additional month's heavy advertising. All of these campaigns used full page advertisements, both in

color and in black and white. Repetition as such, we find, does not insure a cumulative result.

General Conclusions.

Indeed, these tests provided plentiful evidence that all the many factors in advertising, hitherto considered important, such as size, repetition, frequency, position, use of color or black and white, etc., were *relatively* unimportant. For example, a campaign for a toothpaste, appearing entirely in two column advertisements, always in black and white, registered almost three times as well as did a campaign for a rival toothpaste, appearing in full page advertisements, in colored pages, in more magazines, and for a longer time. A single advertisement in black and white was found to have registered its product better than a year's campaign for a rival product, using beautifully colored pages and a wide range of magazines.

Many examples of this kind tend to prove that the one most important factor in successful advertising is the character of the central idea or theme. If this theme coincides with the interests, wants, habits, or feelings of a sufficient number of people, then it will impress its product on their minds. Six weeks after Lux flake soap first began advertising: "Use Lux to cut down runs in stockings," our test on that theme resulted in 47.3 per cent of all the women answering Lux. Runs in stockings is a common irritation among most women, and most women wash their own stockings. Hence a vital and immediate association was assured the use of this theme.

Here again our studies support the conclusions reached by other psychologists, especially by Poffenberger in his penetrating analysis. "There is one requisite for the formation of effective associations that underlies all the others; namely, interest," says Poffenberger. "It is implied, to be sure, in our definition of association which says, that to be associated, two things or ideas must be 'attended to together.' Without interest which is implied

in this phrase, none of the other devices will produce results. Learning without interest and intention is known as 'Incidental Memory' and the surprising thing is the small amount of it. The mechanical devices for directing the course of attention are effective only when interest is present. They are aids, especially when the interest is relatively weak. But where there is strong interest, as in wanting to know what is the means offered for satisfying some imperative need, the mechanical devices are relatively unimportant. Logical relationships then take the place of the mechanical ones." (*Ibid.*, p. 501.)

B. The Psychological Sales Barometer

No matter how correct the conclusion of psychologists that the essential mechanism of advertising is so to establish the name of a product, by association, in the mind of a person that it will be "the one out of all possible trade names that shall come to mind" when a purchase is being made, objective proof of this conclusion has been lacking. Therefore it became desirable to discover facts about the brands being purchased during the period of these tests. The form of question used to obtain this information was generally as follows: What brand of coffee did you buy last? Or, what cold remedy have you purchased or used most recently? The facts found in regard to purchasing habits, as compared with the results of tests on the advertising themes, are illustrated by the following tables:

TABLE IX

Chase & Sanborn Coffee

	March, 1932	Sept., 1932	Oct., 1932	Dec., 1932	Feb., 1933	Oct., 1933
Purchased last by this per cent of women		11.6	12.7		15.6	16.99
Test question answered correctly by	69.1	66.8	71.	75.	75.6 74.1	

Lux Flake Soap

Purchased last by this per cent of women		24.2	24.4	28.6	
Test question answered correctly by	47.7	53.6	70.9		67.5

These two examples represent the longest series of advertising tests and sales barometers in our collection of tests to date. They represent interesting and significant results. Confirming results of lesser scope, however, are plentiful. For example, two tests on the theme: What soap advertises: "Good for B. O. or body odor," gave these results:

Feb., 1933, Lifebuoy, correct answer, by 93.2 per cent of 685 women.

May, 1933, Lifebuoy, correct answer, by 93.1 per cent of 1234 women.

The Lifebuoy soap used by these women was:

April, 1933, of 1348 women, 9.6 per cent used Lifebuoy; of their husbands, 22.3 per cent used it.

October, 1933, of 1932 women, 10.56 per cent used Lifebuoy; of their husbands, 26.04 per cent used it.

We would expect some lag between the establishing of the desired association and the actual use of a product, and this would be particularly true with a soap like Lifebuoy. Moreover, in the field of toilet soaps several other brands, notably Palmolive, Lux, and Ivory, had been already well established in use, particularly among women.

The relationship between effective advertising, as measured by our tests, and the use of the product is by no means a simple causal relationship. For example:

TABLE X

Maxwell House Coffee

	March, 1932	Sept., 1932	Oct., 1932	Feb., 1933	Oct., 1933
Maxwell House purchased last by this per cent of women		13.1	17.4	19.4	21.64
Test question answered correctly by	37.5			42.7	

The increase in the users of Maxwell House coffee was greater than the increase for Chase and Sanborn, even though the advertising was not as effective as that of Chase and Sanborn. The disturbing factor here was probably a radical reduction in price initiated by Maxwell House and only partly followed by Chase and Sanborn.

More complete relationships between advertising and buying habits will be established as these studies continue. In the meanwhile, the development of statistically reliable measures of people's changing buying habits is itself a project of major importance. As an illustration of the results already achieved, the following table is illuminating:

TABLE XI

Barometer—Coffee Brands

Brand	1932 Sept. 3-6 Per Cent	1932 Oct. 3-10 Per Cent	1933 Feb. 17-25 Per Cent	1933 Oct. 2-6 Per Cent
A	13.1	17.4	19.4	21.64
B	11.6	12.7	15.6	16.99
C	7.3	6.5	6.1	5.77
D	5.1	5.8	6.2	5.69
E	4.8	4.8	4.9	5.08
F	4.1	3.9	3.7	3.49
G	2.8	2.8	2.9	3.95
H	2.2	1.7	1.5	.97
I	1.5	1.4	1.3	1.01
J	1.8	1.6	1.7	1.52
Miscellaneous (211 brands)	38.1	35.1	32.6	30.68
None purchased	7.6	6.4	4.2	3.2
Totals	100.	100.1	100.1	99.9
Total sample	541	1445	1399	1932

We have here reasonably consistent results from samples of various sizes, and without any attempt to correct these results for the variations in sectional buying habits which we know exist. Having obtained such consistent

results by a chance distribution of samples, we are justified in believing that a controlled sampling, with proper weighting for geographical sections, will give us a more accurate measure. This is illustrated by the following table:

TABLE XII

Psychological Sales Barometer for Shaving Creams

Based on the study made November 6-10, 1933, in 2386 homes in 40 cities and towns.

Brand	New York State	Eastern States	Southern States	Mid-West	Far West	Weighted Average
Lather Creams						
A	13.5	10.8	12.6	16.4	15.9	13.5
B	11.8	11.6	17.3	9.3	15.3	11.7
C	13.5	12.5	12.0	6.8	12.2	10.5
D	2.7	3.5	6.5	5.4	4.2	4.5
etc.						
Brushless Creams						
E	10.8	12.7	2.9	6.5	6.4	8.7
etc.						
Soaps						
F	3.2	5.5	9.2	9.9	8.5	7.6
etc.						
Miscellaneous	11.3	10.6	10.7	13.1	10.6	11.1
Total per cent	100.	100.	100.	100.	100.	100.
Total sample	461	895	527	728	236	2386

The weighted average is obtained by multiplying each regional percentage by the percentage of the total population represented by that region. The percentage for the Eastern States is multiplied by .46; that for the Mid-West States by .34; that for the Southern and Western States by .10. The sum of the results gives us a weighted or national percentage. A further illustration of this sampling and its use is the following (Table XIII):

TABLE XIII

Psychological Sales Barometer for Cigarettes

Based on the study of November 6-10, 1933, in 2386 homes in 40 cities and towns.

MEN

Brand Smoked	New York State	Eastern States	Southern States	Mid-West	Far West	Weighted Total
A	30.3	30.7	39.7	29.3	26.9	30.73
B	30.3	31.1	23.6	26.2	26.2	28.19
C	24.5	21.6	25.0	28.4	29.6	25.06
D	4.78	4.77	3.76	5.95	7.6	5.39
E	1.91	2.22	.68	1.9	2.07	1.96
etc., etc.						
Total per cent	100.	100.	100.	100.	100.	100.
None, cigars, pipe	147	308	235	308	91	942
Per cent of total men	31.9	34.4	44.6	42.3	39.	38.56
Total sample	461	895	527	728	236	2386

WOMEN

B	43.2	43.1	18.8	29.0	28.9	34.43
C	19.8	19.6	24.6	27.8	30.0	23.93
A	17.5	16.5	39.3	25.7	20.0	22.26
D	2.7	3.3	5.7	5.39	6.67	4.59
E	3.1	4.3	4.9	4.15	6.67	4.54
etc., etc.						
Total per cent	100.	100.	100.	100.	100.	100.
None	239	503	405	487	146	1541
Per cent of total women	51.1	56.3	74.9	66.8	61.8	64.58
Total sample	461	895	527	728	236	2386

These tables show significant differences between brands, territories, and the habits of men and women. It is interesting to note that the ladies who do not smoke are most frequently the ladies of the South.

In order to obtain reliable and significant facts in regard to purchases, it has been found necessary to make our questions more and more specific. For example, the

question: What flake soap did you buy last, which was used in three studies, gave fairly consistent results, but the results were not sufficiently indicative to permit comparison with tests of advertising campaigns. Therefore, we now ask three separate questions (not necessarily all in the same study) namely:

What soap did you buy last for fine washing?
What soap did you buy last for bulk or family washing.
What soap did you buy last for washing dishes?

This gives us both flake and bar soaps, but in terms of three important different uses.

Twenty-four products and about 1500 different brands are now being studied quarterly in this fashion.

We call these studies Psychological Sales Barometers because our measuring rod is in every case a group of people and their purchasing behavior. We are measuring neither the volume of sales nor the sales in dollars, pounds, or units, but *sales trends* as indicated by the changing percentages of brand purchasers. The changes in these percentages from one study to another represent *behavior trends,* the increase or decrease in the number of people who have bought a particular brand. Therefore these barometers are valuable as a measure of all those influences, including advertising, price, and quality of product, which have been brought to bear on the buying public. We cannot always attribute an increase in the proportion of buyers to any one influence, like advertising. However, if our advertising tests show a consistent increase in effectiveness while other influences, including those exercised by competitors, remain the same, we may conclude at least that the advertising is a factor in obtaining new purchasers.

What we are aiming at is not an absolute scientific measure of all factors in the selling and buying situation, but a set of indicators or barometers. The value of such

measures has been well stated by N. M. Balchin, of the National Institute of Industrial Psychology, London, in his recent article: "A Psychological Approach to Market Research." (*The Human Factor*, November, 1933, Volume VII, No. II, p. 375.) Balchin says: "I feel very strongly that market research should be to the manufacturer, not a thermometer, but one of the old-fashioned dialed barometers. Few laymen can deduce much about the weather from reading a thermometer. One can see if it is hot or cold at the moment, but that is all. But the barometer is to some extent prophetic. The glass falls, the pointer swings round to stormy, and the wise man will take warning and go out in his raincoat. The market thermometer of mere statistics helps the manufacturer very little. It records a certain state of affairs without comment. But the market barometer of constructive market research swings its pointer and indicates the wise course of future action."

Commercial Barometer Reports.

In view of the statistical reliability of these barometer reports, arrived at after almost two years of experimental work with methods including actual pantry and bathroom inventories, they are now being issued on a commercial basis. In this way it is hoped to defray some of the heavy expense which these studies increasingly entail, and above all, to pay at least a small sum to the students who have hitherto done the field work without remuneration. Twenty-four products and about 480 different brands are being reported every three months. That is to say, a barometer report similar to *men's shaving creams* illustrated above is made quarterly and a similar report on 23 other products is issued quarterly. Some of the largest companies in the country have already subscribed for one or more series of barometer reports.

Studies of Public Opinion.

Several attempts have been made, in these studies, to evaluate questions as a means of obtaining reliable measures of public opinion in respect to topics of current interest. Questions such as the following were, therefore, introduced from time to time:

Has the idea, "Buy America," led you to refuse to buy any article not made in America?

Do you believe in the farm allotment plan?

What important man, in a political campaign, used the expression "The Forgotten Man"?

Do you think that the United States should cancel foreign debts?

Have you or any member of your family tried Jig-Saw puzzles during the last few weeks?

Do you believe that your city, state, or the Federal Government should take care of unemployment relief?

From what you have seen of the N. R. A. in your neighborhood, do you believe it is working well?

The great increase in recent years in the use of questionnaires, not only by manufacturers but by social agencies of all kinds, makes this subject one of major importance. Our primary object was to discover what types of questions could be relied on to give reliable results. In the question regarding unemployment relief, for example, four alternate answers may be given. Will such a question on the subject of general and secondary interest result in answers representing more than a chance distribution? The answers in this case indicated a rather positive trend of opinion and one which, interestingly enough, is contrary to the trend of events in recent months. These answers were as follows:

TABLE XIV

Who Should Take Care of Unemployment Relief?

February, 1933—39 cities

	Per Cent
City	37.4
State	19.9
Federal Government	27.4
All three	6.0
Don't know	15.5

A more significant set of facts is represented by the answers in two different studies to the same question, namely:

TABLE XV

The N.R.A.

From what you have seen of the N. R. A. in your neighborhood, do you believe it is working well?

Upper figures represent per cent of answers given by 1932 housewives October 2-6, in 40 cities and towns. Lower figures represent per cent of answers given by 2386 housewives November 6-10, 1933, in the same localities.

	New York State	Eastern States	Southern States	Mid-West	Far West	Weighted Average
Yes	50.	45.	55.8	48.2	49.1	47.56
	35.8	35.4	54.9	43.	45.1	40.92
No	29.	30.4	21.55	24.9	24.8	27.09
	35.8	35.	19.9	28.6	24.5	30.26
Don't know	21.	24.6	22.65	26.9	26.1	25.35
	28.4	29.6	25.2	28.4	30.4	28.82

In the weighted averages for the October and November studies we see a definite decrease in the proportion of women who answered *yes*. Still more interesting is the distribution of this decrease between those who answered *no* and those who answered *I don't know*. We see here a process of public opinion in a stage of transition. This change coincided with a fairly well recognized growth of skepticism in regard to the practicality of at least certain details of the N. R. A. Of particular significance

is the distribution of these answers in various sections of the country. Here, the consistency of the results is noteworthy, because the percentages for the different sections vary so widely from the weighted average. The South, for example, is consistently high in its *yes* answers. The East is consistently more doubtful in each study. The results from this same question in our January study, in 50 cities and towns, will add another valuable set of facts to this series.

An interesting sidelight on this subject comes from a study made by a commercial research agency among 40,000 women in four cities during November. In this survey a question regarding the N. R. A. was asked and while unable to give here the exact question asked, published reports of the study in the *New York Times* and other New York newspapers stated that 92 per cent of the women said that the N. R. A. was working beneficially for citizens of the nation. Obviously, the question asked in this study was worded to prompt an emotional or generalized response rather than a response based on actual experience with or knowledge of the N. R. A. and its local effects. The size of the sample, 40,000 women in four metropolitan cities as compared with 4310 women in 40 cities and towns, cannot possibly explain the great discrepancy in these results. This difference must be attributed almost wholly to the forms of questions used.

From such results it is obvious how much the form of the question has to do with the answers obtained, and how precarious is this matter of measuring public opinion. Certainly this is a field peculiarly within the province of psychology, not only from the standpoint of basic experiments conducted in the past, but from the standpoint of developing more scientific techniques for measuring public opinion as a step in analyzing the factors which affect changes in public opinion and public behavior. So far, the bulk of this kind of research has been conducted by economists, sociologists and com-

mercial research agencies, with little or no recognition of the psychological factors and techniques involved in obtaining reliable measures of people's beliefs and behavior. The Psychological Corporation welcomes suggestions not only from its representatives but from other psychologists in regard to fruitful experiments in this field.

MEASURING THE INTERESTS OF COLLEGE STUDENTS WITH DIFFERENT MAJOR SUBJECTS

KATHARINE VAN TUYL and ALVIN C. EURICH
University of Minnesota

When college students select major subjects they designate thereby very definite interests. Without doubt the choice of a major in English literature represents different interests from the decision to major in physiological chemistry. At present counselors have no technique available for crystallizing the interests of bewildered students prior to the selection of major courses of study. Frequently when asked what he wishes to do after he leaves college, the student responds by saying that he desires to teach. When faced with the question of what he proposes to teach, he is confused. It would be helpful indeed if an interest analysis blank could be given to the student to determine whether his interests are like those of English or science teachers. To aid in the solution of this problem an attempt to construct interest blanks that differentiate between the various major groups has been made at the University of Minnesota. This paper presents a brief evaluation of the results.

General Procedure

The Minnesota Interest Blank was given to juniors and seniors in the College of Education in December 1928. The blank is a revision made by Jacobsen[1] of

[1] Marion Myer Jacobsen. "The Quantitative Determination of Scholastic Interests Among College Students." Unpublished Ph.D. thesis. University of Minnesota. 1928. 165 pp.

the original Freyd questionnaire, adapted to the measurement both of scholastic interests and of social-mechanical interests. The resulting form contains 100 occupational and 100 miscellaneous items; to each there are four possible responses—L for like, D for dislike, O for indifference, and U for unknown.

The records obtained were then subdivided according to the major subjects represented. The large variety of majors made it necessary to combine related fields into single groups; namely, the natural sciences, the social science, and the languages. The groups thus formed and the number of students in each were as follows: educational administration-supervision (15), art (10), English (90), foreign language (46), mathematics (30), music (13), natural science (37), physical education (11), social science (72), total (324).

College ability ratings of the entire sample were obtained from the records of the Miller Analogies Test, the Minnesota Reading Examination, and the Peterson Equation Completion Test, all given in the junior year. The percentile rank of each student on each test was transmuted into a standard score on an arbitrary scale in terms of tenths of standard deviation units with the mean at 50, and zero at 5 S.D.'s below the mean. An average of the standard scores of each student in the three tests was then obtained. These average scores are henceforth referred to as ratings in "intelligence."

Among the larger groups the variation in mental ability was estimated from the significance of the differences between the measures of central tendency and variability. The groups were then equated in intelligence by elimination of as many individuals as necessary to obtain for each group a mean of 50. No attempt was made to equalize differences in variability for the necessary elimination would have rendered the groups too small to be used in the comparisons planned.

Interest scores were first obtained according to the Freyd Scale for social-mechanical interest and the Jacobsen Scale for scholastic interests. The mean scores were used to indicate the trends of the various groups.

A Comparison of Major Group Interests

Each major group was compared with every other group according to individual responses to the interest blank. In the groups used the numbers of individuals of each sex, after reduction to hold intelligence constant, were as follows:

Major Subject	Number of Cases		
	Men	Women	Total
English	6	72	78
Social Science	19	51	70
Foreign Language	0	32	32
Natural Science	18	19	37
Mathematics	1	13	14
Administration	13	1	14
Total	57	188	245

The items diagnostic of the above groups were derived first by converting the number of responses to each item for every group into a percentage of the total number of cases. For each item the difference between the percentages of each two groups giving the same response was then analyzed to determine whether or not it was "significant." To calculate the degree of significance the ratio of the difference to its probable error was obtained. The procedure may be made clear by the following illustration:

Per Cent of English Group	Per Cent of Social Science Group	D	PE_D	$\frac{D}{PE_D}$
26	44	18	.05	3.6

A difference of 18 per cent was found between the English and the social science majors checking "0" opposite the item "advertising expert." The difference divided by its probable error of .05 gave a ratio of 3.6. According to the criterion of 3PE used in this study, the difference of .18 was selected as "significant." The stimulus "advertising expert" is thus a diagnostic item which differentiates the English and the social science groups.

By this method the items distinguishing two groups were selected. For each item in the questionnaire the difference between the responses to each of the four symbols, L, D, O, and U, and the statistical significance of that difference were determined. The Edgerton-Paterson Table of Standard and Probable Errors of Percentages was used for these computations.

The numbers of "significant differences" between each two groups were found to be as follows:

Natural Science	vs.	Mathematics	91
		Administration	106
		Foreign Language	166
		Social Science	171
		English	276
Mathematics	vs.	Administration	77
		Foreign Language	123
		Social Science	163
		English	168
Administration	vs.	Foreign Language	163
		Social Science	170
		English	224
Foreign Language	vs.	Social Science	125
		English	106
Social Science	vs.	English	120

These values were used to indicate the relative similarities and dissimilarities between groups in their interests. Thus the mathematics majors are most like those in educational administration while the natural science and English groups diverge most widely in their likes and dislikes. In general the English, social science, and language groups are comparatively similar in interests and differ from the science, mathematics, and administration groups which are also somewhat similar.

Sex Differences

In order to estimate the influence of sex differences, the responses differential between men and women were first isolated in the natural science and social science groups, the only two groups in which the sexes were sufficiently comparable in number. The 81 diagnostic items found in the natural science group and the 119 found in the social science group were combined into a list used as a tentative representation of the responses affected by sex differences. These responses were selected out of the total number of differential items between certain of the major groups. For example, out of the 91 items distinguishing between the natural science and mathematics groups significant sex differences were found in 31 or 34 per cent. While the elimination of these items reduced the total number of significant differences in each case, in no instance did it affect the relative similarities and dissimilarities between the groups studied. The natural science, mathematics, and administration groups were still comparatively alike in interests, as opposed to the language, social science, and English students which were similar to each other.

It was considered permissible therefore to combine the more similar groups and thus form larger groups in which a more thorough study of sex differences could be made. The natural science, mathematics, and admin-

istration groups were combined into a single group which may be called the "science-mathematics" group, and the language, social science, and English groups formed a "social-language" group. These two large divisions included respectively 65 and 180 cases. By a comparison of the total number of responses to each item, 243 significant items were isolated that constituted a scale by which the total sample of 324 cases was scored. Thus every student obtained a score according to the interests of the general group to which he belonged.

To derive a separate scoring scheme for one sex the 33 women in the science-mathematics group were compared with the 155 women in the social-language group. The analysis gave 174 significant responses.

The two scoring schemes, that for women and that for both sexes, were compared to determine what items they might have in common, and what items might show by disagreement between the two keys apparent sex differences. The "common items" proved to be 155 in number, or 59 per cent of the 243 items differential between the entire groups. The remaining 41 per cent consisted of items either affected to varying degrees by the elimination of one sex, or affected merely by the reduction in the number of cases. It was observed that the value of $\frac{D}{PE_D}$ frequently changed but very little by the use of only one sex. The number of items in the women's scoring scheme was reduced by the elimination of numerous differences whose ratios to their probable errors were less than 3.00, but which varied by only a few decimal points from those obtained when both sexes were used. Since larger probable errors are obtained with a smaller number of cases, changes resulting from such slight variations in ratios may be assumed due to the reduction in the size of the groups. Among all the non-identical items the amount of difference between the two keys ranged from a slight change in the

degree of difference to an actual reversal of the direction of the difference.

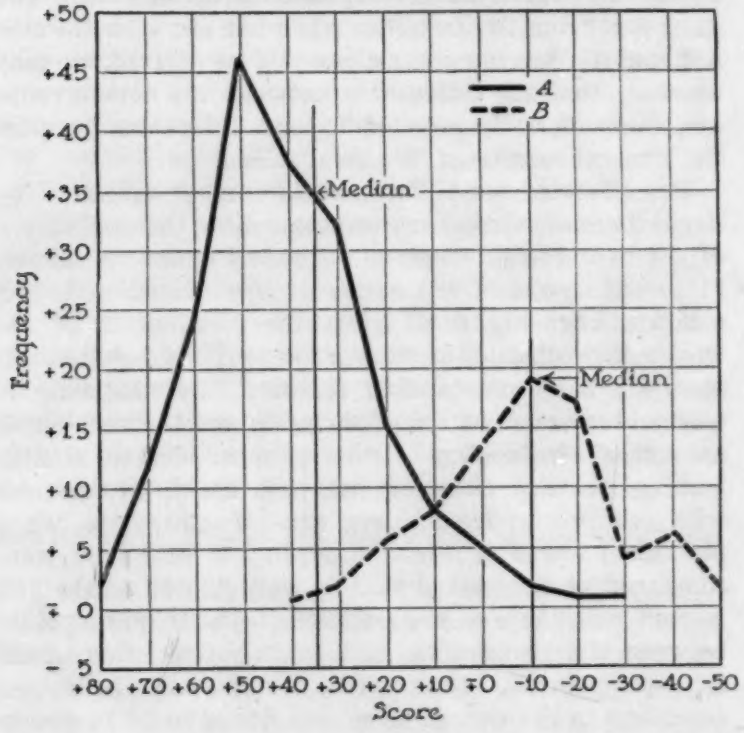
Since 59 per cent of the items remain unchanged when women only are used, and since the other 41 per cent vary greatly in the degree and probable significance of the changes which have occurred, it is evident that sex differences have affected the responses to a certain extent, but to what extent cannot be exactly determined. From the degree of correspondence between the two keys, and from the fact that when one sex alone is used a diagnostic scoring scheme can still be derived, we may conclude that the differences between the total groups are not sufficiently affected by sex differences to alter the general results of the comparison.

The effect of sex differences is further indicated by the individual interest scores obtained by the application of the two scoring schemes. Figure I shows the distribution of scores of the entire groups according to the common key; Figure II gives the distribution of the women according to the women's key. Both distributions show the same outstanding features; the clustering of scores about one portion of the scale and the very small amount of overlapping between groups. Closely similar results are thus obtained with the entire groups and with groups restricted to one sex. Furthermore, when individual scores of women according to both keys were compared, it was found that in only 33 out of the 188 women was there a discrepancy of 15 or more points between the two scores, and in only five cases of 20 or more points. The quartile deviation of the scores according to the women's key was found to be 12 points and that of the scores according to the common key 19 points. A range of 15 points is therefore slightly more than one quartile deviation on one scale and slightly less than one quartile deviation on the other scale. The remaining 155 cases showed discrepancies of less than 15 points.

This comparison of individual scores and their distributions, together with the comparison of the items composing the two scales, led to the conclusion that the differences between groups were related primarily to

FIGURE I

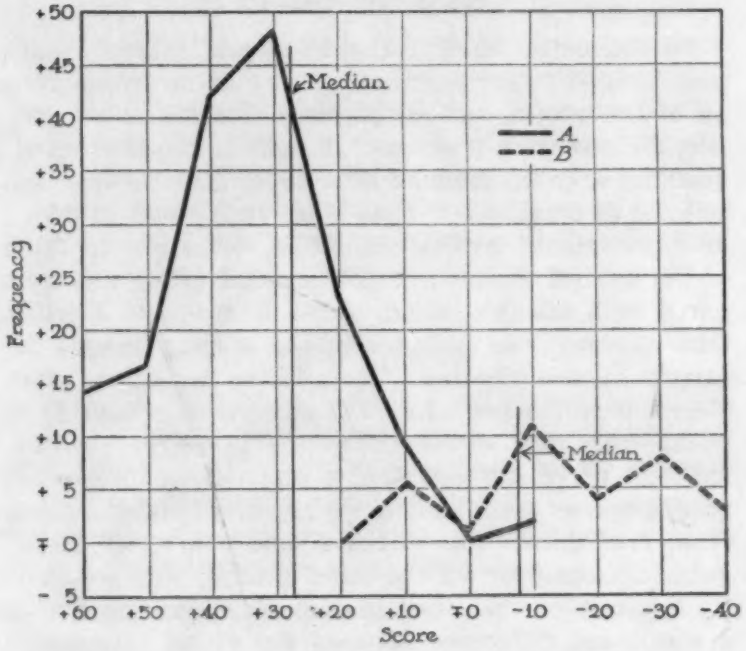
Distribution of Social-Language (A) and Science-Mathematics (B) Groups According to Interest Scores



major interests rather than to the interests of one or the other sex, and that the derivation of diagnostic scales for major groups might be attempted with the inclusion of both sexes.

FIGURE II

Distribution of Women of Social-Language (A) and Science-Mathematics (B) Groups According to Interest Scores



The results thus far obtained show also that the interests of two general classes of students may be differentiated. For vocational guidance it is of more importance, however, to determine whether the interests of students specializing in certain specific subjects such as English or science may be distinguished.

The Derivation of Diagnostic Interest Scales for Major Groups

To determine whether the Minnesota Interest Blank may be used to differentiate any one major group from all of the others, various methods of selection of diagnostic items were tried out. Finally a response distinguishing a group from at least three other groups was selected as diagnostic. Thus if the response *L* to item 1 or "accountant" gave a significant difference in favor of the natural science group when that group was compared with administration, social science, and English, that response was included in the scoring scheme for natural science interests. The selected items were given weighted scores indicating the number of groups from which each item distinguished the group in question. An item which differentiated a group from three other groups was scored 1, an item which differentiated a group from four others was scored 2, and one which distinguished from 5 or all the other groups, was scored 3. By this method each item in each scale was based upon a significant difference between the group in question and three or more other groups; each score assigned indicated the degree to which the item was differential of that group. The sum of the weighted scores according to any scale constituted a total individual score.

By this procedure the following numbers of diagnostic items were selected for the various groups: natural science 37, mathematics 20, administration 34, foreign language 23, social science 24, and English 36.

The Application of Interest Scales to Major Groups

The scales derived in this manner were applied both to criterion and to control groups. To obtain control groups the questionnaire was given to juniors and seniors in November 1931, and the records scored according to the techniques based upon the criterion groups. The groups thus obtained were as follows: administration 18, English 57, foreign language 29, mathematics 24, natural science 45, social science 63, total 236. In these control groups no attempt was made to hold intelligence constant.

In both the criterion and control groups each individual was scored according to each scale. Thus an English student received a score not only in English interests, but in mathematics interests, language interests, etc.

By the application of the interest scales which had been derived it was aimed to determine their reliability, validity, and value for the prediction of scholarship.

Reliability

Since the method of retests was not feasible at the time the interest blanks were given, the reliability of the major interest scales was determined by the odd-even technique. The coefficients estimated by the Spearman-Brown formula were as follows, the number of cases in each group being given in parenthesis:

	Criterion Groups		Control Groups	
Mathematics	(14)	.37	(24)	.30
Administration	(14)	.80	(18)	.86
Foreign Language	(32)	.69	(29)	.42
Natural Science	(37)	.60	(45)	.49
Social Science	(70)	.67	(63)	.79
English	(78)	.74	(57)	.80

In view of these results it may be stated that the administration, English, and social science scales show the greatest reliability; the natural science and foreign

language scoring schemes are less reliable, and mathematics the least reliable. On account of the irregularities in the results further tests of reliability are needed.

VALIDITY

The validity of the scales was determined by their effectiveness in differentiating the major group, or by the overlapping among groups on interest scores. From the scores of each individual on each scale—the scores in mathematics interests, language interests, etc.—the percentage of each group reaching or exceeding the median of every other group was calculated. The separation of each group from other students in general was also determined; i. e., the percentage of all non-language students treated as a single group reaching the median on the language scale was computed. The amount of overlapping was determined in this manner for the criterion and for the control groups, and also, in order to estimate the differentiation which might be obtained with larger groups, for the combined criterion and control groups.

Table I shows the overlapping among criterion groups on the foreign language scale. The overlapping of other majors on the language group ranges from zero to 16.6 per cent, and of language majors on other groups from zero to 18.7 per cent. Seven and five-tenths per cent of all non-language students reach or exceed the median of the language scale. In Table 2 the differentiation between control groups on the English interest scale is given. The separation obtained between the combined criterion and control groups on the natural science scale is indicated in Table 3. Similar calculations were made for criterion, control, and combined groups on each of the interest scales.

From the overlapping among all groups used it was found that the most successful differentiation was secured

TABLE I

Differentiation of Foreign Language Criterion Group from Other Criterion Groups

Overlapping of Other Groups on Foreign Language						Overlapping of Foreign Language on Other Groups				
Group	N	Range	Median	Cases Reaching Median FL	Percentage Reaching Median FL	Group	Range FL	Median FL	Cases Reaching Median	Percentage Reaching Median
Natural Science	37	5-23	10.0	0	0	Natural Science	5-33	13.0	0	0
Mathematics	14	6-26	17.0	1	7.1	Mathematics	1-15	6.0	0	0
Administration	14	5-22	10.5	0	0	Administration	3-29	10.0	0	0
Social Science	70	2-28	17.0	2	2.9	Social Science	6-31	17.0	6	18.7
English	78	7-32	21.0	13	16.6	English	14.55	34.5	4	12.5
All Non-Foreign Language Students	213	---	---	16	7.5					

with the natural science scale, on which the total overlapping in criterion, control, and combined groups ranged from 0.5 to 6.8 per cent. The mathematics scale is next in degree of diagnostic ability, with a total overlapping from zero to 7.5 per cent. The overlapping of all other students on the foreign language scale ranged from 7.5 to 16.4 per cent and that on the English scale from 7.8 to 19.5 per cent; these two scoring schemes occupied a middle ground with respect to differentiating capacity. The social science group was less differentiated than any of the foregoing; from 17.7 to 23.1 per cent of other students reached the median of the social science scale. The diagnostic ability of the administration scale was uncertain since the total overlapping of criterion groups was only 0.9 per cent, while for the control groups it was 24.3 per cent. This scoring scheme appeared to be least valid when applied to control groups.

The factors in the greater overlapping on the social science and administration scales are uncertain. These students may be less distinct in their interests, or the Interest Blank may be less adapted to the measurement of their likes and dislikes. If the former is true, a more rigid method of scoring is needed to isolate their interests.

The Relation of Major Interests to Scholarship

To study the relation between interest scores and achievement in the major subject rank order correlations were obtained between honor point ratios in the major subject and interest scores. The resulting coefficients were as follows:

Administration	—.13±.17
Mathematics	—.38±.11
Foreign Language	—.03±.14
Natural Science	—.02±.11
English	+.41±.08
Social Science	+.13±.09

TABLE II

Differentiation of English Control Group from Other Control Groups

Overlapping of Other Groups on English						Overlapping of English on Other Groups				
Group	N	Range	Median	Cases Reaching Median E	Percentage Reaching Median E	Group	Range E	Median E	Cases Reaching Median	Percentage Reaching Median
Natural Science	45	4-42	21.0	4	9.1	Natural Science	2-35	13.0	2	3.5
Mathematics	24	3-42	20.0	1	4.2	Mathematics	1-16	6.0	1	1.8
Administration	18	7-47	31.5	5	27.8	Administration	1-32	9.0	2	3.5
Foreign Language	29	12-54	37.0	10	34.5	Foreign Language	6-30	19.0	15	26.3
Social Science	63	10-55	32.0	15	23.8	Social Science	8-32	19.0	17	29.8
All Non-English Students	179	---	---	35	19.5					

TABLE III

Differentiation of Total Natural Science Group from Other Control Groups

Overlapping of Other Groups on Natural Science						Overlapping of Natural Science on Other Groups				
Group	N	Range	Median	Cases Reaching Median NS	Percentage Reaching Median NS	Group	Range NS	Median NS	Cases Reaching Median	Percentage Reaching Median
Mathematics	38	12-40	22.0	6	15.8	Mathematics	3-16	7.0	7	8.5
Administration	32	7-43	24.5	6	18.8	Administration	6-33	16.0	6	7.3
Foreign Language	61	5-33	14.0	0	0	Foreign Language	2-27	11.0	4	4.9
Social Science	133	0-37	14.0	4	3.0	Social Science	2-28	11.0	3	3.7
English	135	1-36	14.0	3	2.2	English	1-42	21.0	0	0
All Non-Natural Science Students	399	---	---	19	4.8					

In the wide variation shown by these results, the outstanding tendency is toward a zero correlation. In the first four groups the wide range of the probable errors, together with the unreliability of the natural science and language scales, make the results very uncertain. All that may be said regarding the possible degree of correlation between interests and scholarship in these four groups is that the obtained coefficients may be either positive or negative and will probably be very small. Furthermore, the administration and mathematics groups were found to be quite highly selected with respect to scholarship, with a concentration in the intermediate levels. This fact adds to the unreliability of the results so that there is little evidence that a negative correlation of any size would be found in a random selection of these two major groups. The English and social science groups yield the most reliable coefficients which are, moreover, based upon the most reliable of the scales.

Summary and Conclusions

Briefly, the study purported to explore the possibility of measuring objectively the unique interests of students who have selected different major courses of study. After the Minnesota Interest Blank was given to college juniors and seniors, the items yielding significant differences between the responses of any two major groups were selected. On the basis of the number of differentiating items, the "science-mathematics" and the "social-language" groups were formed since it was found that these two groups differed markedly in interest scores. An analysis of sex differences within these groups indicated that the small differences between men and women were not sufficient to account for the differences in major subject interest.

In the final derivation of the major subject interest scales, each item was weighted according to the number

of groups from which it distinguished the group in question. By this means, interest scales were derived for English, social science, foreign language, natural science, mathematics, and educational administration groups. The results secured through the application of these scales to other groups warrant the following generalizations:

(1) Interest scales can be constructed that differentiate objectively between various major subject groups.

(2) When major subject interest scales were derived from the Minnesota Interest Blank, the natural science scale proved to be most valid in terms of differentiating the natural science group from other majors. The social science scale is least valid, while the data concerning the educational administration scales are inconsistent.

(3) The reliability of the major subject interest scales as estimated by the Spearman-Brown formula varies from .30 to .86. The educational-administration and English scales appear to be most reliable while the mathematics scale appears to be least reliable.

(4) The varying degrees of differentiation obtained suggest that certain major groups such as natural science possess more distinct interests than other groups.

Now that interest scales have been constructed that differentiate between major subject groups, the next task is to determine whether the scales can be used to predict the unique interests of students before they select their major courses of study. The validity of the scales for this purpose remains undetermined.

MEMORY FOR ADVERTISING COPY PRESENTED VISUALLY VS. ORALLY

FRANK N. STANTON[1]
The Ohio State University

The present experiment was designed to determine the comparative effectiveness of fictitious advertising copy presented to the subjects by the printed page versus that delivered through a radio loudspeaker. The method in many respects follows that used by DeWick.[2]

The advertising copy was presented by loudspeaker without a program in the sense of music, dialogue, etc., while the printed material was placed before the subjects void of illustrations, display types and other attention arresting devices. The effort was to eliminate as far as possible all variables but visual versus auditory presentation.

The advertising copy used for both auditory and visual presentation consisted of two groups of eight advertisements each, hereafter denoted, Copy X and Copy Y. With one group of subjects for instance, Copy X was given visually and the Copy Y by loudspeaker. With another group X was auditory and Y visual. The commodities in each series represented articles frequently purchased, universal in use, and low in unit cost. Included in each list there was one product of especial interest to men and another especially attractive to women.

[1] The author wishes to acknowledge the kind cooperation of Dr. Harold E. Burtt of The Ohio State University who contributed useful and constructive criticisms. Dr. F. H. Lumley also of The Ohio State University assisted by helpful advice.

[2] DeWick, Henry N., "The Recall Value of Visual and Auditory Advertising Material" (unpublished).

The remaining six products were presumably purchased equally by both sexes.

The fictitious trade names selected for the 16 commodities in the two lists were two syllable names chosen on the basis of a preliminary study with visual and auditory presentation. Confusing final consonants such as s and f were distributed equally in the two lists. No two names ended with the same syllable or syllables that sounded identically. The same precaution was maintained for the first syllable of each word. The two lists were balanced for similar products wherever necessary, *e. g.*, Gasoline *vs.* Oil, Tea *vs.* Coffee. The two lists of trade names and their products are printed below.

X SERIES	Y SERIES
Lorenz Powder	Barlow Cigarettes
Harvey Toothpaste	Parker Coffee
Bryan Gasoline	Crawford Bread
Warren Candy	Stewart Oil
Kenwood Soap	Robbins Shaving Cream
Douglas Cigars	Andrews Gingerale
Curtis Ink	Browning Hosiery
Randolph Tea	Wagner Mouthwash

A further attempt was made in the preliminary series to insure equal trade-name-commodity-associations. A group of 40 subjects, similar to those used in the experiment proper, participated. The experimenter read a list of fictitious trade names to the group with the instructions to write the name of any product they could associate with the trade name. After collecting those responses the experimenter repeated the fictitious trade names and with each one gave a product which he had arbitrarily selected to go with it. To all subjects this presentation was made orally. A portion of them received typed copies of the fictitious trade name commodity lists which they read as the experimenter presented the list verbally. Immediate recall and aided

recall tests were given. After an analysis of both sets of responses, those trade names for which there existed any consistent product associations other than the fictitious ones of the experiment were eliminated. Likewise cases where the fictitious trade-name-commodity-associations ran extremely high or low with the experimental group were eliminated. The effort was to select material for which there was no naturally strong association between product and brand and also in which the formation of such associations was not extremely difficult.

Each piece of advertising copy was of approximately constant length—between 70 and 75 words—and included trade name product mentions three times. The first mention occurred at the opening, another approximately in the middle with the third at the close of the paragraph. In each mention the trade name always came first with the product immediately after. Care was exercised to keep all copy as near equal in complexity and sales appeal as possible. No dramatic statements or slogans were employed. A sample piece of copy is presented below.

"For better times, Andrews Gingerale. A pleasant dinner, an enjoyable evening and then the grand finale. Only the sparkle and zest of this famous pale dry Andrews Gingerale will make your party complete. The new large bottles hold enough for five full glasses, making anywhere from five to eight long, tall drinks. And the last glass out of the bottle is just as good as the first. Andrews Gingerale holds its life!"

There were 8 such pieces in the X and Y sets—16 in all.

Subjects for this investigation were 160 students of both sexes enrolled in psychology courses at The Ohio State University during the summer quarter of 1933. They were divided into four groups to permit the necessary reversals of copy and method of presentation. Table I shows how copy and method of presentation varied with the 4 groups. For instance, Group A had

Copy X auditory followed by Copy Y visual at the same sitting.

TABLE I

Groups	A	B	C	D
Copy Used and Method of Presentation	X aud	Y aud	X vis	Y vis
	Y vis	X vis	Y aud	X aud
Number of Subjects	39	38	42	41

The visual presentation was effected by the use of booklets. These measured 6 inches in width and 9 inches in height and used an antique India[3] stock for the text with a cover of dark blue paper. The instructions and fictitious advertising copy were set in 12 point, bold, antique face,[4] 18 picas, leaded two points.[5] No emphasis was given to trade names or products by the use of caps, italics or special type. Copy only occurred on every fourth page making it necessary to use 40 pages in each booklet. The cover bore the word "Front" set in 48 point Goudy bold with a double rule under the word on the X series and a single rule under "Front" on the Y series.

The same copy as prepared[6] for the booklets was used

[3] India Mayflower Text, 60 lb. basis, Central Ohio Paper Co., Columbus, Ohio.

[4] 12 pt. bold antique, designed for use with DeVinne, light face, Mergenthaler Linotype Company, Brooklyn, N. Y.

[5] Printing specifications followed in general the findings of Paterson, Donald G., and Tinker, Miles A., "Studies of Typographical Factors Influencing Speed of Reading": VIII. "Space Between Lines or Leading." *Journal of Applied Psychology*, 1932, 16, 388-397.

[6] An electrical transcription of the auditory presentation was prepared on a Fairchild Professional Record. This was not used in the experiment.

for the auditory presentation. For this purpose a small portable public address system and a student announcer were used. The public address system had a maximum output of 7 watts. With it were used a double-button microphone and a dynamic loudspeaker. The announcer read the copy to the groups. Several presentations were made with the announcer and experimenter for practice in order to eliminate possible stress or emphasis on any point. The trade names were not spelled out. The announcer and amplifying unit were outside the classrooms in which the auditory presentations were made.

The following directions were given by the experimenter when the visual presentations came first (Group C and D, Table I).

"Shortly, I shall pass out some booklets to you. Please place them so that the word 'Front' on the upper half of the cover is facing you.[7] When I give the signal you will please break the seal and turn at once to the first printed page. After you have finished reading it please turn to the next page which will be blank and then look up to receive the starting signal. Ready . . . . begin."

Further instructions for the visual part of the experiment were presented on the same paper with the same type as the advertising copy in order to facilitate adjustment of the subjects for the copy to come. The instructions read by the subjects were as follows:

"On the following pages you will read a series of advertisements for familiar products. Will you please sit comfortably and read carefully the copy. Later[8] you will be asked to answer some

[7] Each copy of the booklet had a 1½" band of white paper wrapped around it upon which these instructions were printed: NOTICE: Do not open this booklet until you have been instructed to do so.

[8] Realising that some subjects would suspect subsequent testing for the content of the ads we felt it would be wise to spread their attention as much as possible over all the copy by the general statement noted.

questions about the products and the trade names—the features of the products and the inducements to buy. Time enough will be allowed to read each advertisement through once. When you have finished reading the ad please turn immediately to the next page in this booklet. The page will be blank. After you have done so await the signal from the experimenter before turning to the next advertisement. The time between readings will give you an opportunity to think over the ad just read. Throughout the experiment will you please remain quiet so as not to disturb your neighbor."

Before giving the signal to begin the experimenter gave an opportunity for questions but none were asked in the entire experiment. Following that he gave the starting signal. At the end of each succeeding minute the signal to turn was given until all eight advertisements were read. The booklets were collected and after a pause came the following announcement through the loudspeaker which had been installed before the class convened:

"In the following minutes you will hear a series of advertisements for familiar products. Will you please sit comfortably and listen carefully to the copy. Later you will be asked to answer some questions about the products and the trade names—the features of the products and the inducements to buy. Between each advertisement there will be a pause during which you will have time to think over the ad just heard. Throughout the experiment will you please remain quiet so that you will not disturb your neighbor?"

After those instructions the experimenter, still present, asked for questions and when the class was ready he so signalled the announcer who began to deliver the auditory series over the public address system. The time consumed in delivering each piece of copy ran approximately 35 seconds. Auditory presentations began every 60 seconds thereby allowing a 25 second pause between each delivery. Three seconds before the announcer began with the copy proper he said "ready" in an effort to set the subjects. This corresponded to the set produced

in the visual series by the signal "turn". With Groups A and B the same method was followed except that auditory presentations came first.

Recall and recognition tests were given at three different time intervals after the presentations, 1 day, 7 days, and 21 days. Three subgroups of subjects were formed for this purpose. A subgroup was made up by selecting every third name from the class rolls. On the test day the instructors read the names of the students who were to serve as subjects. They left their classes and met in a common room. Thus each group was tested only once. In our determinations of effectiveness of the two media, we were concerned only in correct responses to trade-name-commodity-associations built up by both methods of presentation. For each subgroup the following tests were given in the order named: Pure Recall, Aided Recall, and Recognition.

The Pure Recall Test forms were mimeographed and headed by the following directions to the subjects:

> "On the lower part of this sheet will you please list as quickly as possible all the products you can remember as having heard or read about in this experiment. List them in the order in which you recall them. Try to list both product and trade name, such as: Beechnut Chewing Gum, not just Chewing Gum. If you can not do that, then please list either the product or trade name in the proper column. Please try to do your very best. Keep your eyes on your own sheet at all times. A sample is given below."

For the 1 day and 7 day groups approximately 20 minutes were required before everyone finished and the blanks were collected. The 21 day groups required about 10 minutes on the same form.

The Aided Recall Test forms were given as soon as all Recall blanks had been collected. Two forms of the Aided Recall Tests were designed in order to keep the commodity names on the blank in the same sequence as that in which they had been in the booklet and through

the loudspeaker or *vice versa*. This test form consisted of a series of 16 blank lines with one commodity appearing to the right of each line. Only products actually used in the experiment were included in the list. The instructions on each form of the test were as follows:

"On the lower part of this sheet you will find a list of products made up from the advertisements you heard and read in the experiment. In front of each product on the list is a blank space. Write in the blank space the trade name for each product on the list that you can remember. Please keep your eyes on your own paper at all times. A sample is given below."

Not over 10 minutes were used by any test group in satisfying the conditions of this form.

The Recognition Test blanks were made up likewise and consisted of 16 blank lines each of which was followed on the right by the product and four, two syllable, fictitious trade names. One out of each four was correct and was to be recorded in the blank. The positions of the correct trade names were staggered down the page. Directions for this test follow:

"On the lower part of this sheet you will find a list of products made up from the advertisements you heard and read in the experiment. In front of each product is a blank space in which you are to write the correct trade name of each product if you recognize the trade name in the group of four trade names appearing to the right of each product listed. Please keep your eyes on your own sheet at all times. A sample is given below."

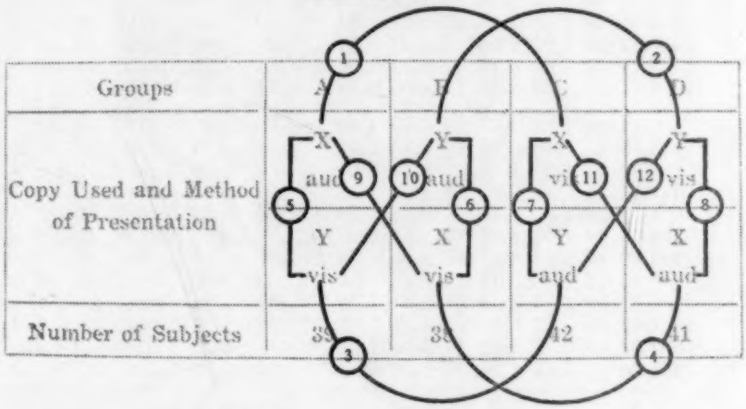
After the completion of the three test forms, the groups were asked by the experimenter not to inform the remaining subjects of the nature of the testing. Obviously the remaining subjects knew of the testing and that in all probability they too would be used.

The individual test forms were scored and tables, including the means, standard deviations and analyses of significant differences were built up for each group on each of the three test periods for the three types of tests.

From Table I it can readily be seen we are prepared to make several comparisons between auditory and visual scores. To facilitate the visualization of these comparisons or breakdowns Table II has been prepared. It is identical to Table I except for the connecting lines whose encircled numbers label the comparisons.

In our analysis of the data collected in this experiment we shall make all the possible auditory-visual comparisons that will contribute worthwhile information. For instance: Comparison 1 involves different subjects, one having Copy X auditory as its first presentation and the other having Copy X visual as its first:

TABLE II



Comparison 5 involves the same subjects taking Copy X auditory followed by Copy Y visual.

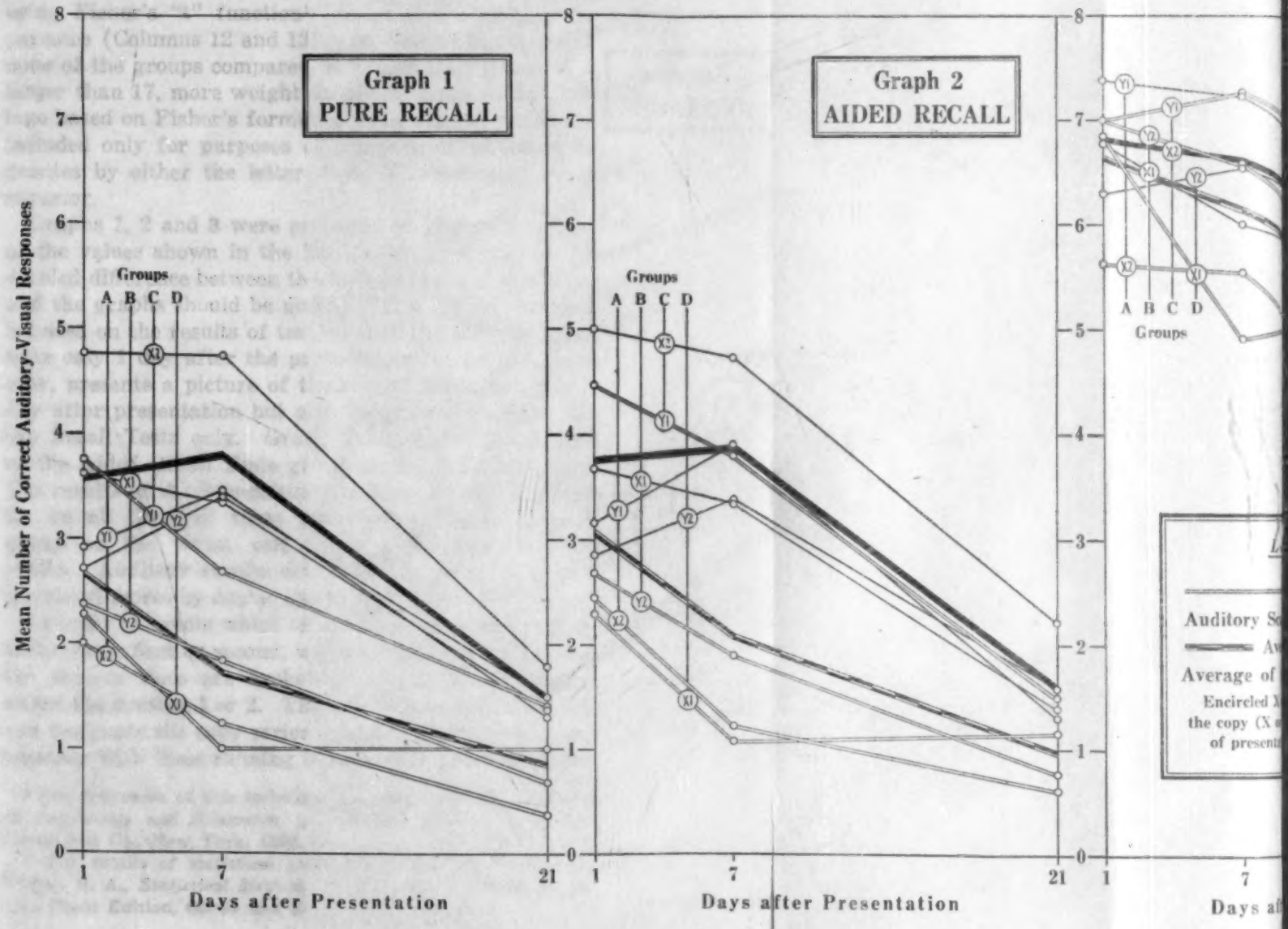
Table III shows how the first 53 subjects responded to the three tests 1 day after presentation. Table IV and V similarly show the differences for other subjects for the two media 7 days and 21 days after presentation respectively. In the three tables each of the 12 comparisons (Table II) is presented showing both auditory and visual mean scores, (Columns 2, 5), their differences, (Column

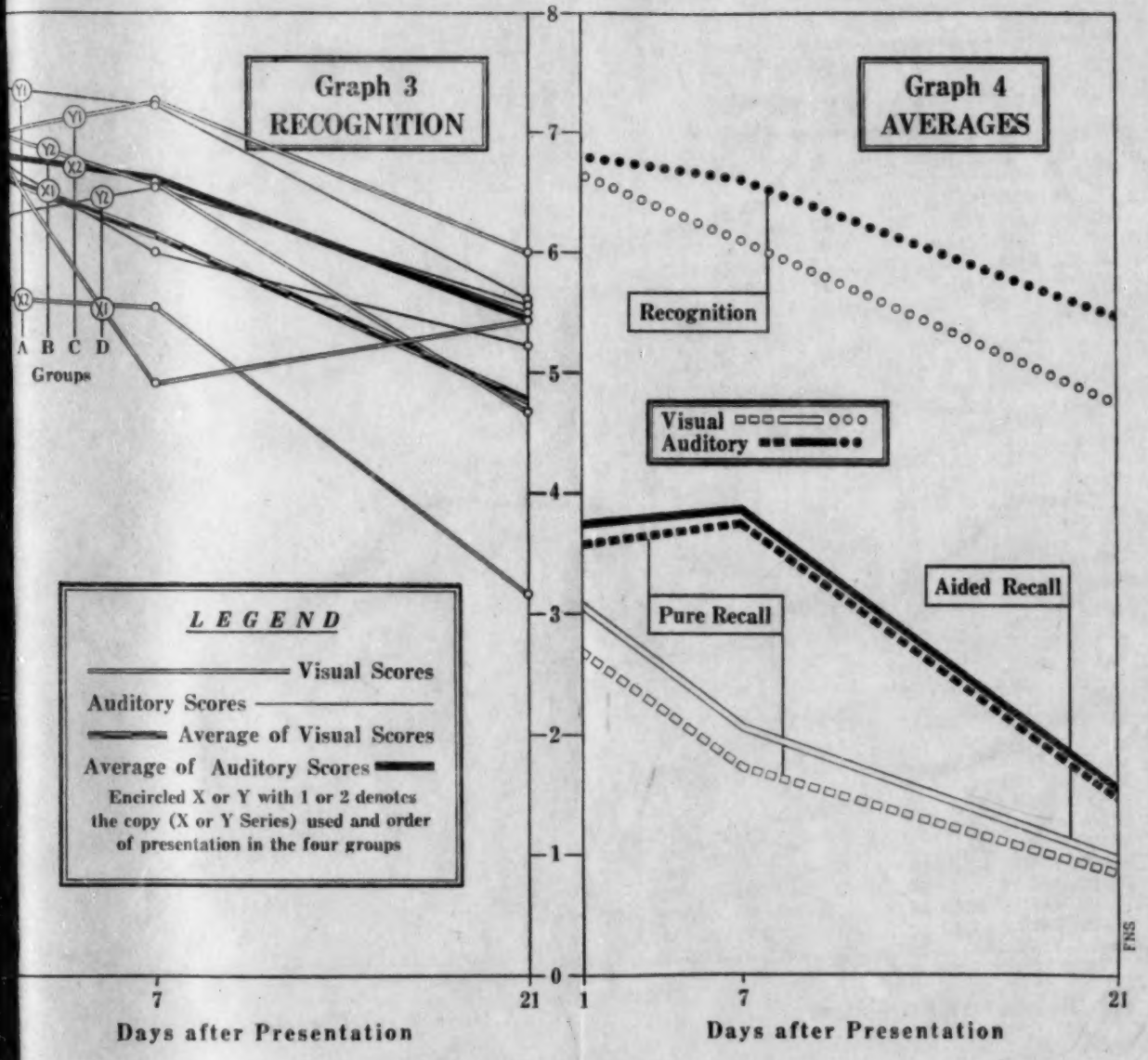
TABLE III

Results: 1 Day after Presentation

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
	Comparison	Auditory Mean	Auditory N	Auditory σ	Visual Mean	Visual N	Visual σ	Difference	r	σ diff.	Diff/σ diff	t	P	Superior
Recall	1	2.92	13	1.73	3.76	17	2.31	.84		.74	1.14	1.0598	.30-.20	V
	2	3.61	13	1.98	2.28	14	1.90	1.33		.75	1.77	1.7146	.10-.05	A
	3	4.76	17	2.29	2.00	13	1.41	2.76		.67	4.12	3.6977	.01*	A
	4	2.71	14	1.83	2.38	13	1.44	.33		.71	.05	.4984	.70-.60	A
	5	2.92	13	1.73	2.00	13	1.41	.92	.52	.43	2.14	2.0004	.10-.05	A
	6	3.61	13	1.98	2.38	13	1.44	1.23	.21	.61	2.01	1.9156	.10-.05	A
	7	4.76	17	2.29	3.76	17	2.31	1.00	.40	.60	1.67	1.5313	.20-.10	A
	8	2.71	14	1.83	2.28	14	1.90	.43	.36	.62	.69	1.8456	.10-.05	A
	9	2.92	13	1.73	2.38	13	1.44	.54		.62	.87	.8406	.50-.40	A
	10	3.61	13	1.98	2.00	13	1.41	1.61		.67	2.40	2.3232	.05-.02	A
	11	2.71	14	1.83	3.76	17	2.31	1.05		.81	1.29	1.3353	.20-.10	V
	12	4.76	17	2.29	2.28	14	1.90	2.48		.75	3.30	3.1302	.01*	A
Aided Recall	1	3.15	13	2.89	4.47	17	2.64	1.32		1.02	1.29	1.2579	.30-.20	V
	2	3.69	13	1.73	2.28	14	2.37	1.41		.79	1.78	1.6887	.20-.10	A
	3	5.00	17	2.17	2.46	13	2.37	2.54		.85	2.99	2.9481	.01*	A
	4	2.86	14	2.38	2.69	13	1.53	.17		.76	.22	.2103	.90-.80	A
	5	3.15	13	2.89	2.46	13	2.37	.69	.60	.66	1.04	1.3487	.30-.20	A
	6	3.69	13	1.73	2.69	13	1.53	1.00	.42	.49	2.04	1.8416	.10-.05	A
	7	5.00	17	2.17	4.47	17	2.64	.53	.38	.66	.80	.7383	.50-.40	A
	8	2.86	14	2.38	2.28	14	2.37	.58	.47	.65	.89	.8631	.50-.40	A
	9	3.15	13	2.89	2.69	13	1.53	.46		.90	.51	.4942	.70-.60	A
	10	3.69	13	1.73	2.46	13	2.37	1.23		.82	1.50	1.4713	.20-.10	A
	11	2.86	14	2.38	4.47	17	2.64	1.61		.80	2.01	1.7083	.10-.05	V
	12	5.00	17	2.17	2.28	14	2.37	2.72		.82	3.32	3.2227	.01*	A
Recognition	1	7.38	13	.62	7.00	17	1.14	.38		.33	1.15	1.0497	.40-.30	A
	2	6.77	13	.96	6.77	14	1.20	.00		.00	.00			O
	3	6.82	17	1.34	5.61	13	1.73	1.21		.57	2.12	2.0865	.05-.02	A
	4	6.29	14	1.71	7.00	13	.96	.71		.53	1.34	1.2681	.30-.20	V
	5	7.38	13	.62	5.61	13	1.73	1.77	.69	.38	4.66	4.6838	.01*	A
	6	6.77	13	.96	7.00	13	.96	.23	.66	.12	1.92	.9975	.40-.30	V
	7	6.82	17	1.34	7.00	17	1.14	.18	.02	.42	.43	.4028	.70-.60	V
	8	6.29	14	1.71	6.77	14	1.20	.48	.33	.46	1.04	1.0474	.40-.30	V
	9	7.38	13	.62	7.00	13	.96	.38		.32	1.19	1.1654	.20-.10	A
	10	6.77	13	.96	5.61	13	1.73	1.16		.55	2.11	2.0565	.10-.05	A
	11	6.29	14	1.71	7.00	17	1.14	.71		.54	1.31	1.3353	.20-.10	V
	12	6.82	17	1.34	6.77	14	1.20	.05		.45	.01	.1044	1.0-.90	A

* P is less than .01


Graph 1
PURE RECALL
Graph 2
AIDED RECALL
Mean Number of Correct Auditory-Visual Responses
Groups
A B C D
Days after Presentation
Auditory Sc
Av
Average of
Encircled X
the copy (X a
of present


Graph 3
RECOGNITION
Y1
Y1
Y2
X2
X1
Y2
X2
X1
A B C D
Groups
LEGEND
Visual Scores
Auditory Scores
Average of Visual Scores
Average of Auditory Scores
Encircled X or Y with 1 or 2 denotes the copy (X or Y Series) used and order of presentation in the four groups
7
21
Days after Presentation
Graph 4
AVERAGES
Recognition
Visual
Auditory
Pure Recall
Aided Recall
8
7
6
5
4
3
2
1
0
1
7
21
Days after Presentation
FNS

8), the significance of these differences as determined by the conventional critical ratio[9] (Column 11), and also by using Fisher's "t" function[10] designed for small comparisons (Columns 12 and 13). In view of the fact that none of the groups compared in Tables III, IV and V is larger than 17, more weight should be given to the findings based on Fisher's formulae. The critical ratios are included only for purposes of comparison. Column 14 denotes by either the letter A or V which medium is superior.

Graphs 1, 2 and 3 were prepared to illustrate certain of the values shown in the Tables III, IV and V. One decided difference between the treatment given the tables and the graphs should be noted. Table III for instance is based on the results of testing with the three types of tests only 1 day after the presentation. Graph 1, however, presents a picture of the results taken not only 1 day after presentation but also 7 days and 21 days, for the Recall Tests only. Graph 2 illustrates the results of the Aided Recall Tests given on the three test days. The results of the Recognition Tests are shown in Graph 3. In all three of these graphs the results of each group on the visual and auditory presentations are shown. Auditory results are shown by single lines and the visual scores by double lines.

In order to denote which of the two methods of presentation came first or second, without referring to Table I, the various lines are broken by circles which enclose either the number 1 or 2. The letters X and Y in the circles designate the copy series used. The groups are tied together with lines running between the pairs of circles

[9] For discussion of this technique see Garrett, H. E., *Statistics in Psychology and Education*, pp. 128-134, 286-288, Longmans. Green and Co., New York, 1926.

[10] For details of statistical technique of the "t" function see Fisher, R. A., *Statistical Methods for Research Workers*, pp. 99-111, Third Edition, Oliver and Boyd, Edinburgh, 1930.

in order to aid in showing the group trends. Each pair is marked by its group letter. In addition to these lighter lines on each graph there are two bold lines. These represent the average of all results for each medium on the particular test form shown in the graph. More will be said of the heavier lines later.

An inspection of Table III shows that for pure recall one day after the presentations the auditory method is superior in 10 out of the 12 comparisons (Table II). This decided plurality for the auditory medium is probably significant in itself but statistically only Comparisons 3 and 12 may be regarded as genuine differences in the light of the critical ratio and the "t" function.

The 12 comparisons on the Aided Recall Tests produce about the same picture statistically. It is worthy of note that throughout the experiment the results of the Aided Recall Tests follow very closely the Recall Test scores. However, when the Recognition Test scores are considered, the difference between the auditory and visual method shrinks. At 24 hours after presentation only half of the comparisons favor audition.

There may be several explanations for this change. The fact that the visual responses seem to approach the auditory in recognition may be due to the fact that 8 is a perfect score and both media are approaching their upper limits. Again, it is argued that one might expect inferior visual recognition of copy presented orally. If the latter be true, then the results shown in Table III for auditory recognition after 24 hours should be given more weight. Poffenberger[11] points out that recall is perhaps a better check on memory where there are no attention factors operating while to make a high recognition score implies higher attention value of the advertisement. The copy used in this experiment employed

[11] Poffenberger, Albert T., *Psychology in Advertising*, page 192, Second Edition, McGraw-Hill, New York, 1932.

TABLE IV

Results: 7 Days after Presentation

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
	Comparison	Auditory Mean	Auditory N	Auditory σ	Visual Mean	Visual N	Visual σ	Difference	r	σ diff.	Diff/σ diff	t	P	Superior
Recall	1	3.46	13	1.08	2.67	15	2.14	.79		.63	1.25	1.1597	.30-.20	A
	2	3.33	12	1.79	1.00	13	.78	2.33		.56	4.16	4.0984	.01*	A
	3	4.73	15	2.62	1.23	13	.70	3.50		.70	5.00	4.5057	.01*	A
	4	3.38	13	1.44	1.83	12	2.19	1.55		.66	2.34	2.0210	.10-.05	A
	5	3.46	13	1.08	1.23	13	.70	2.23	.69	.22	10.13	8.6500	.01*	A
	6	3.33	12	1.79	1.83	12	2.19	1.50	.64	.44	3.40	2.4110	.05-.02	A
	7	4.73	15	2.62	2.67	15	2.14	2.06	.80	.40	5.15	5.5966	.01*	A
	8	3.38	13	1.44	1.00	13	.78	2.38	.60	.32	7.43	7.1287	.01*	A
	9	3.46	13	1.08	1.83	12	2.19	1.63		.60	2.71	2.2894	.05-.02	A
	10	3.33	12	1.79	1.23	13	.70	2.10		.55	3.82	3.7605	.01*	A
	11	3.38	13	1.44	2.67	15	2.14	.71		.68	1.04	.9767	.40-.30	A
	12	4.73	15	2.62	1.00	13	.78	3.73		.71	5.25	4.7653	.01*	A
Aided Recall	1	3.92	13	1.38	3.80	15	2.68	.12		.79	.15	.1399	.90-.80	A
	2	3.33	12	1.79	1.08	13	.73	2.25		.56	4.02	4.0049	.01*	A
	3	4.73	15	2.62	1.23	13	.97	3.50		.73	4.79	4.3874	.01*	A
	4	3.38	13	1.59	1.92	12	2.13	1.46		.75	1.95	1.8725	.10-.05	A
	5	3.92	13	1.38	1.23	13	.97	2.69	.69	.28	9.61	8.6087	.01*	A
	6	3.33	12	1.79	1.92	12	2.13	1.41	.62	.50	2.82	4.2375	.01*	A
	7	4.73	15	2.62	3.80	15	2.68	.93	.93	.26	3.58	4.4864	.01*	A
	8	3.38	13	1.59	1.08	13	.73	2.30	.80	.40	5.75	7.5530	.01*	A
	9	3.92	13	1.38	1.92	12	2.13	2.00		.72	2.77	2.6955	.02-.01	A
	10	3.33	12	1.79	1.23	13	.97	2.10		.58	3.62	3.5389	.01*	A
	11	3.38	13	1.59	3.80	15	2.68	.42		.82	.51	.4761	.70-.60	V
	12	4.73	15	2.62	1.08	13	.73	3.65		.71	5.10	4.6846	.01*	A
Recognition	1	7.23	13	.80	7.27	15	1.44	.04		.43	.09	.0860	1.0-.90	V
	2	6.00	12	1.87	4.92	13	1.54	1.08		.69	1.56	1.5167	.20-.10	A
	3	6.60	15	1.70	5.54	13	1.33	1.06		.57	1.86	1.7512	.10-.05	A
	4	6.54	13	1.82	6.58	12	1.38	.04		.64	.06	.0587	1.0-.90	V
	5	7.23	13	.80	5.54	13	1.33	1.69	.17	.40	4.22	3.9308	.01*	A
	6	6.00	12	1.87	6.58	12	1.38	.58	.62	.43	1.35	1.7263	.20-.10	V
	7	6.60	15	1.70	7.27	15	1.44	.67	.78	.28	2.39	2.4896	.05-.02	V
	8	6.54	13	1.82	4.92	13	1.54	1.62	.23	.58	2.79	2.6836	.02-.01	A
	9	7.23	13	.80	6.58	12	1.38	.65		.46	1.41	1.3957	.20-.10	A
	10	6.00	12	1.87	5.54	13	1.33	.46		.65	.71	.6840	.60-.50	A
	11	6.54	13	1.82	7.27	15	1.44	.73		.62	1.18	1.1405	.30-.20	V
	12	6.60	15	1.70	4.92	13	1.54	1.68		.61	2.75	2.6240	.02-.01	A

* P is less than .01

TABLE V

Results: 21 Days after Presentation

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
	Comparison	Auditory Mean	Auditory N	Auditory σ	Visual Mean	Visual N	Visual σ	Difference	r	σ diff.	Diff/σ diff	t	P	Superior
Recall	1	1.30	13	1.20	1.40	10	1.91	.10		.68	.15	.1470	.90-.80	V
	2	1.38	13	1.15	1.00	14	1.25	.38		.46	.83	.7891	.50-.40	A
	3	1.80	10	1.47	.38	13	.49	1.42		.48	2.96	3.1113	.01*	A
	4	1.50	14	1.34	.69	13	1.32	.81		.52	1.56	1.5212	.20-.10	A
	5	1.30	13	1.20	.38	13	.49	.92	.03	.35	2.63	2.0004	.10-.05	A
	6	1.38	13	1.15	.69	13	1.32	.69	.46	.36	1.92	3.3020	.01*	A
	7	1.80	10	1.47	1.40	10	1.91	.40	.75	.40	1.00	1.3094	.30-.20	A
	8	1.50	14	1.34	1.00	14	1.25	.50	.48	.35	1.43	1.6138	.20-.10	A
	9	1.30	13	1.20	.69	13	1.32	.61		.49	1.24	1.1995	.30-.20	A
	10	1.38	13	1.15	.38	13	.49	1.00		.34	2.94	2.8023	.01*	A
	11	1.50	14	1.34	1.40	10	1.91	.10		.69	1.45	.1439	.90-.80	A
	12	1.80	10	1.47	1.00	14	1.25	.80		.58	1.38	1.3755	.20-.10	A
Aided Recall	1	1.30	13	1.20	1.50	10	2.46	.20		.85	.23	.2450	.90-.80	V
	2	1.38	13	1.15	1.14	14	1.50	.24		.51	.47	.4465	.70-.60	A
	3	2.20	10	1.89	.61	13	.62	1.59		.61	2.60	2.7138	.02-.01	A
	4	1.57	14	.90	.77	13	1.58	.80		.50	1.60	1.5706	.20-.10	A
	5	1.30	13	1.20	.61	13	.62	.69	.47	.29	2.38	2.4073	.05-.02	A
	6	1.38	13	1.15	.77	13	1.58	.61	.46	.41	1.49	2.5804	.05-.02	A
	7	2.20	10	1.89	1.50	10	2.46	.70	.50	.70	1.00	1.6560	.20-.10	A
	8	1.57	14	.90	1.14	14	1.50	.43	.45	.36	1.19	1.2557	.30-.20	A
	9	1.30	13	1.20	.77	13	1.58	.53		.55	.96	.9366	.40-.30	A
	10	1.38	13	1.15	.61	13	.62	.77		.36	2.14	2.0666	.05-.02	A
	11	1.57	14	.90	1.50	10	2.46	.07		.82	.08	.0940	1.0-.90	A
	12	2.20	10	1.89	1.14	14	1.50	1.06		.70	1.51	1.4672	.20-.10	A
Recognition	1	5.61	13	2.12	6.00	10	1.79	.39		.82	.48	.4465	.70-.60	V
	2	5.23	13	1.37	5.43	14	1.29	.20		.51	.39	.3764	.80-.70	V
	3	5.50	10	1.96	3.15	13	1.99	2.35		.83	2.83	2.6996	.02-.01	A
	4	5.57	14	1.59	4.69	13	1.81	.88		.65	1.35	1.2941	.30-.20	A
	5	5.61	13	2.12	3.15	13	1.99	2.46	.58	.52	4.73	5.1727	.01*	A
	6	5.23	13	1.37	4.69	13	1.81	.54	.59	.41	1.32	1.1710	.30-.20	A
	7	5.50	10	1.96	6.00	10	1.79	.50	.44	.60	.83	.6680	.60-.50	V
	8	5.57	14	1.59	5.43	14	1.29	.14	.37	.43	.32	.3076	.80-.70	A
	9	5.61	13	2.12	4.69	13	1.81	.92		.77	1.19	1.1566	.30-.20	A
	10	5.23	13	1.37	3.15	13	1.99	2.08		.67	3.06	3.0184	.01*	A
	11	5.57	14	1.59	6.00	10	1.79	.43		.70	.61	.5937	.60-.50	V
	12	5.50	10	1.96	5.43	14	1.29	.07		.70	.01	.1008	1.0-.90	A

* P is less than .01

no attention getting devices voluntarily. Consequently, on that basis we would expect a shrinkage in the differences between the two media.

Table IV presents the data for the scores made 7 days after the presentation. The differences at this point in the testing are the greatest of the three testing dates for the recall tests. All 12 comparisons on the Recall Test scores favor the auditory method. Eleven out of the 12 Aided Recall Test differences are in the direction of audition. Seven of the 12 auditory differences for recall are clearly significant. Nine out of the 11 of the differences favoring audition on the Aided Recall Tests likewise are statistically significant. Recognition Test scores taken at this period show a slight superiority in favor of audition with 7 out of 12 in that direction. One of the 7 differences is significant according to the data.

The data secured from the tests given 21 days after the presentation show a decided fall in the amount remembered according to all three test forms. Material delivered by the auditory method still leads on all test forms but the difference is very slight in many cases. The Recall Test results at 21 days show 11 out of 12 comparisons in favor of audition. Three of the 11 differences are statistically significant. The results of the Aided Recall Tests are very close with 11 out of 12 on the side of audition. Here one of the differences is statistically significant. Auditory recognition results are strongest at the 21 day period for 8 out of the 12 differences are in that direction, three of them being definitely significant. The increasing difference between the auditory and visual recognition results is clearly shown in Graph 3 by the constant increasing divergence between the average lines for the two media of the test periods.

Up to this point the data have been presented by group breakdowns just as the material was presented. This has been done in order to give a fair picture of the results because certain differences and significant varia-

tions are oftentimes obscured when scores obtained under slightly different conditions are grouped and averages computed. However, to get an idea of the general trend we have grouped all the auditory scores and all the visual scores for each test on the three test days. The grouping involves all the auditory and visual scores of Comparisons 5, 6, 7 and 8. Out of the 12 possible auditory visual comparisons (Table II) it seems reasonable that the comparisons on the same subjects 5, 6, 7 and 8 should be given more weight. The reversals of the methods of presentation together with the changes in the order of copy reveal no extreme differences due to these positional factors. We have assumed that the effects of these variables have been balanced out in the averages, *e. g.*, Copy X occurred first as many times as second and the visual method of presentation was used first as many times as second.

The results of this grouping are shown in Table VI. In this table the data are given for the three test periods and for the three types of tests. Differences have been computed between the auditory and visual averages for each test form on each test period of the grouped scores. From the differences the critical ratios have been calculated and are also presented in Table VI. The auditory method of presentation is superior on each test day for all three types of tests. Statistically, only the comparisons between audition and vision for the Recall Tests 1 day, 7 days and 21 days after presentation and the comparisons for Aided Recall Tests 7 days and 21 days after presentation are significant.

Graph 4 illustrates the comparisons made in Table VI. It is identical to the bold or average lines shown in Graphs 1, 2 and 3. From this graph it is very evident that the responses made to the Aided Recall Tests are just a little higher than those to the Recall Tests. The auditory trend is indicated by solid black and the visual by outline. The greatest difference between the auditory

TABLE VI

		1 Day		7 Days		21 Days	
		Aud	Vis	Aud	Vis	Aud	Vis
Recall	Mean	3.58	2.68	3.77	1.72	1.48	.84
	σ	2.16	2.10	1.97	1.76	1.15	1.35
	Difference	.90		2.05		.64	
	r	.47		.81		.65	
	σ diff	.29		.16		.15	
	Diff/σ diff	3.10		12.81		4.27	
	Superior	Aud		Aud		Aud	
Aided Recall	Mean	3.75	3.07	3.89	2.08	1.58	.98
	σ	2.21	2.48	2.02	2.17	1.33	1.45
	Difference	.68		1.81		.60	
	r	.51		.83		.69	
	σ diff	.31		.17		.15	
	Diff/σ diff	2.19		10.65		4.00	
	Superior	Aud		Aud		Aud	
Recognition	Mean	6.81	6.63	6.60	6.11	5.48	4.76
	σ	1.30	1.40	1.66	1.38	1.70	2.02
	Difference	.18		.49		.72	
	r	.18		.38		.40	
	σ diff	.22		.24		.29	
	Diff/σ diff	.82		2.04		2.48	
	Superior	Aud		Aud		Aud	
	N	57		53		50	

and visual method of presentation is shown for the Recall Tests at the 7 day period. The slight rise in the auditory recall curve at the 7 day testing is an interesting variation to the usual performance of a curve of retention. It may be due to the fact that the tests were given to different subgroups on each test date. Yet that alone can not be given as an explanation for if we plot the data from series I of DeWick's[12] study where the same subgroups were tested on each test date a parallel rise in

[12] Op. cit., page 45.

the auditory recall curve is noted while the visual trend is also parallel to the curve shown in Graph 4 of this report.

The approximate identity of the Recall and Aided Recall curves can also be clearly seen in Graph 4. Recognition for the two methods of presentation seems to be very nearly equal after a lapse of one day, but as the time goes on the auditory method increases its lead over the visual technique. That is contrary to the results obtained for the Recall and Aided Recall testings for in both the peak difference occurred at 7 days. It should be noted that we obtained data at only 3 points. Consequently, it is quite possible that the actual peak difference might occur somewhat earlier or later than 7 days.

The results of the experiment agree within limits with the work done by DeWick. He employed only a recall test which was given to all subjects immediately, 1 day, 7, and 150 days after stimulation. Immediate recall showed very little difference in favor of the auditory medium. This difference increased up to 5 days in Series I and up to 7 days in Series II.

Chiefly because the results would have few practical applications, and secondly due to the equality of results DeWick secured, the immediate test period was omitted in the present study. The 1 day, 7 day and 21 day periods were selected arbitrarily.

An examination of Graph 4 will give some indication of the magnitude of the correct responses for each test form. Auditory recognition scores run well ahead of the Aided Recall and Recall mean scores. Such results are not contrary to the usual comparisons of this nature. Of the three test forms, perhaps the Aided Recall test is nearest to answering questions of a practical nature that are raised by the advertiser when considering which of the two media he will select to carry his story. If he is advertising products similar to the ones used in this study, his chief desire is to have the consumer think of

his brand or trade name when a need arises for that product.

Throughout the experiment every precaution was taken to control all the variables by keeping them as constant as possible. Yet in searching the procedure and results for an answer to the reason for auditory superiority, one must not overlook the part played by the experimental situation. The auditory results are probably heightened somewhat as a result of the abnormal attention to the loudspeaker and the copy. But one must not overlook the visual advantages. Few printed advertisements receive the attention given to the visual copy in this study. The students were warned not to look inside the booklets, thus setting up a condition of expectancy. They were instructed to read every word of copy which, of course, is not the case in normal visual advertising. Thus the effectiveness of both media presumably was stepped up. Here, as in the case of presenting the copy (visual void of illustrations, display faces, etc.—auditory without music or sound effects), the advantages and disadvantages are as nearly equal as possible with this set-up. We must not overlook the participants who were college students, and their surroundings, habits, ages and interests. The writer plans to continue this means of measurement on listeners in home situations. When we consider that the college student is a trained reader, such an experiment with other persons may even show a greater difference in favor of audition for certain economic levels.

SUMMARY

Fictitious advertising copy was presented visually by means of printed booklets and orally through a small public address system to four groups of college students. Testing 1 day, 7 days and 21 days after stimulation by Recall, Aided Recall and Recognition tests for correct

trade-name-commodity-association built up by the two media showed the auditory method superior. Not all cases showing an auditory superiority were statistically significant. The peak of the superiority in the Recall Tests came for the 7 day period. At the 21 day check both auditory and visual results were much lower and the margin between the two had suffered by the lapse of time. Throughout the experiment little difference in scores was noted between the Aided Recall and Pure Recall tests. In the recognition scores, the greatest auditory superiority occurred at 21 days. The principal variable that might indicate the desirability of qualifying the results was the discrepancy arising from the experimental situation between actual reading-listening conditions and the ones set up in the study.

A SUMMARY AND CRITIQUE OF METHODS OF MEASURING "WARMING-UP"

E. B. SKAGGS
Wayne University

Up to the year 1924 many psychologists were inclined to look upon warming-up, whenever found, as a manifestation of improvement due to ordinary practice or learning. Other psychologists were inclined to regard WU[1] as something over and above learning as considered in the usual sense. In 1924 Robinson and Heron[2] devised a simple formula for ruling out improvement due to practice. As it is the purpose of this paper to give a brief summary and critique of present and past methods of measuring WU and as the Robinson-Heron method is probably the most used at present, we shall consider this formula first of all.

THE ROBINSON-HERON METHOD

Robinson and Heron reported two experiments, one involving tapping and one involving the recitation of the alphabet backwards. Applying their formula they found definite WU in both tasks when a certain discontinuity of work was involved. The formula which these investigators used, expressed in percentage terms, may be stated in either of two ways, as follows:

$$\frac{100(y-x)}{x} - \frac{100(z-x)}{x} > 0,$$

or,

$$\frac{100(y-x)}{x} > \frac{100(z-x)}{x}$$

[1] Hereafter we shall use WU to signify "warming-up."

[2] E. S. Robinson and W. T. Heron. *Journal of Experimental Psychology*, 1924, 7, 81-97.

In each of the above formulae,

x stands for the total number of taps during the first 5 minutes of the day's work, i. e., the first half of the work;

y represents the total number of taps during the second 5 minutes of the first day's work, i. e., the second half, and

z designates the number of taps during the first 5 minutes of the second day's work, i. e., the first half of the day's work.

These writers wrote as follows concerning their method: "The warming-up may be said to be demonstrated where the increase from a first to a second five minutes of tapping is greater than that from a first five minutes of tapping on one experimental day to the first five minutes of tapping on the immediately succeeding experimental day."[3] In the above quotation the reader will understand that a total day's work consisted of 10 minutes; hence, instead of speaking of the first and second five-minutes one can speak of the first and second half of each day's work.

Unfortunately the method is open to several criticisms. The writers themselves have pointed out one criticism, namely, the fact that the formula is very severe and may credit too much to practice. The writers state the criticism precisely, as follows, "Our test of the presence and degree of warming-up effect, if it is in error, errs on the side of too great severity. The amount of practice is measured by improvement from the first part of one experiment day to the first part of the next. The practice which may be assumed to affect the next day's tapping is that included not only in the first five minutes of one day, but also in the second 5 minutes of that day." In a word, small amounts of WU may be masked by the use of the formula. This is a very serious objection if one wishes to measure WU in very minute amounts.

[3] Robinson and Heron. *Op. Cit.* Page 85.

The method also assumes that no learning or consolidation process takes place after the day's work is ended. This is a precarious assumption because many bits of evidence indicate the fact of a perseveration or consolidation process which continues after active practice ceases. In a word, the first half of the second day's work may be better than the second half of the first day's work or just as good because of the effects of the perseveration processes and not due to the practice factor which operated to make the second half of the first day's work better than the first half of the day's work.

Method of Computing the Average Value of Practice

A method which was used by one of Kraepelin's students, Dr. Emil Amberg,[4] and one which antedates the Robinson-Heron method by many years, attempts to find and utilize the average practice-effect over a series of periods or days of work. As this method has never been translated into English or discussed, and inasmuch as Kraepelin used and accepted the method, it seems desirable to consider the method in some detail. Amberg used among other tasks continuous addition of two digits. In the case under consideration the work-period lasted for one hour in every case. In the control experiment the subject added continuously for one hour. In the experiment proper the subject worked continuously for a half-hour, rested 5 minutes, and then worked continuously for a second half-hour. The amount of work done in the first and second half-hour in both the control and experimental conditions was computed. By a computation which we shall show directly the average practice-effect for the entire series of eight days was found. Hav-

[4] Emil Amberg. *Uber den Einfluss von Arbeitspausen auf die geistige Leistungs-fahigkeit.* Leipzig, 1895. Also see *Psychologische Arbeiten,* Vol I.

ing found this average practice value, this value was then subtracted from every difference in results between the first and second half of every day's work.

We may make use of some hypothetical figures to make clear the method. Let the Roman numerals below represent the eight consecutive days of work in addition. Let the figures below the Roman numerals represent the units of work done each day. In order to simplify matters the figures are shown in an order of regularly increasing value. Actually, of course, this would not he the case, or at least, this is not the usual situation.

I	II	III	IV	V	VI	VII	VIII
10	15	20	25	30	35	40	45

(1) The value of each work-day, considered as a base, was compared with every succeeding work-day value and the differences were expressed in percentage terms. Thus, the value of I was compared with the values of II, III, IV. . . . . VIII. The value of II was compared in like manner with III, IV . . . . VIII. The value of III was compared with the five succeeding days, and so on, until every value had been compared with every succeeding value. Using our hypothetical figures the following computations would be made and the following figures obtained, all in percentage terms:

I	II	III	IV	V	VI	VII	VIII
50	33	25	20	17	14	13	----
100	66	50	40	33	29	--	
150	100	75	60	50	--		
200	133	100	80	--			
250	166	125	--				
300	200	---					
350	---						
7)1400	6)698	5)375	4)200	3)100	2)43	1)13	
200	116	75	50	34	22	13	

(2) Having found the difference between each work-hour and every other following work-hour, as shown in the columns of figures above, the second step consisted in finding the sum of each list of figures. These sums are shown by the figures 1400, 698, 375, etc.

(3) The average value of each list of percentage values was then computed. These quotients are shown above, as 200, 116, 75, etc.

(4) The quotients to which we have just referred were now summed. In the above case the sum would be 510.

(5) This sum (510) was then divided by 28, the total number of comparisons made. This quotient was considered by Amberg to represent the *average practice value*, expressed in percentage terms. In the above case this would be 18 per cent.

(6) Having obtained this very important value, WU was then computed by comparing the gain of the second half of each day's work over the first half, and subtracting from this gain, if any, the amount to be credited to practice. If, in our example, the increase of the second over the first half-hour on day I was found to be 22 per cent, then, since the average practice-effect was found to be 18 per cent, 4 per cent of this gain would be due to pure WU.

One might use this same method but deal with absolute units of work instead of percentage values. Thus, let us assume that the average practice effect is of the value of 2 units of work. One could then compare the first and last half of any given day's work, making the proper correction for practice. Any increase of the second over the first half would be credited to WU to the extent that it was greater than 2. This may be illustrated as follows:

Day I		Day II		Day III	
First Half	Second Half	First Half	Second Half	First Half	Second Half
Value, 4	Value, 6	Value, 6	Value, 9	Value, 11	Value, 9
Gain, 2		Gain, 3		Loss, 2	
Pract., 2		Pract., 2		Pract., 2	
WU, 0		WU, 1		WU, 0	

If this method of computing and applying the average practice-value be invalid a whole monograph of computations of WU falls accordingly because Amberg devised and used this method throughout his work. Likewise, if this method is incorrect, many of Kraepelin's statements about WU are without adequate basis because he accepted the technique.

One very serious criticism may be made at once. Since the curve of practice is usually negatively accelerated the average practice effect, as computed by Amberg, would become smaller and smaller with the addition of practice-trials and result in some spurious results. Since the practice-effects would actually be large during the very early practice-periods and very small near the end of the series of work-days, the use of the average practice-value, would result in a disproportionately great amount of WU in the first section of practice and a disproportionally small amount of WU in the later sections. If one could work with well-defined segments of the total learning curve and take an average for each segment this criticism might be obviated.

Method of Direct Comparison Where Practice Effects Are Negligible

Ideally, in searching for WU, one should first train a subject until he has reached his limit of improvement in a given task. A search of the experimental literature reveals but few cases where the subjects have been so highly trained that they worked near the limit of their capacity in the given task. It may take months and

even years to reach one's limit of improvement in a given task, especially if the task is a relatively complicated or difficult one. On the other hand it is entirely possible, by an expenditure of patience and time, so to train a subject in the more simple psycho-motor or mental tasks that no further improvement is noted. Having reached this stage the experimenter would be justified in discarding all formulae and compare the first with the later sections of a given day's work. This program only awaits the advent of psychologists who will make the sacrifice of time and energy in order thoroughly to train their subjects!

Method of Curve Inspection

Robinson and Heron[5] present fifteen curves of progress in the recitation of the alphabet backwards. Every one of these curves shows a rather rapid and sharp rise at the beginning of the work-period. The initial rise was much sharper than the general upward trend of the curve for the work done over the entire 60 minutes of work (recitation). These writers note that ". . . . there is a greater improvement during the brief minute or so of this warming-up than occurs from the beginning of one cycle of the experiment to the next—a period containing 60 minutes of actual practice." One may argue, and very legitimately, we believe, for the presence of true WU when the curve shows an initial sharply marked upward trend which is far more steep than the general upward trend of the learning curve, but the fact remains that this is a qualitative rather than a quantitative method. Aside from any criticism that might be made when one argues for WU on the basis of inspection of the curves themselves, the essential weakness of the method lies in the fact that it does not permit of quantification.

[5] Robinson and Heron. *Op. Cit.* Quotations from page 96.

Method of Rank Order Comparison

This method consists in arranging in rank order of value or efficiency the trials or other units, as seconds or minutes, which are being compared. Thus, one would arrange the trials or successive intervals of work-time in order from the most to the least efficient for each day's work. One would then study these rankings to find whether or not there was a consistent trend for an increase of efficiency to occur at a given point in the progress of the work. If the second trial were better than the first in a given task, for example, in 75 per cent of the cases considered, one might argue for WU. Again we are dealing with a method which does not permit of quantitative treatment, just as was the case in the above-mentioned method. Also, the method fails to take into account the effects of practice or learning.

Method of Utilization of Plateaus

A method which has never been used heretofore and one which may prove to be invalid makes direct use of marked plateaus occurring in the course of learning. If one trains a subject in a given task, such as the sorting of cards or the working of the formboard, and continues this training over a long course of time until the subject is highly expert at the task, several marked plateaus are likely to appear. The writer has never failed to obtain less than three to six or more pronounced plateaus as he has trained subjects to work the formboard. These plateaus were flat and of long duration.

The plateau, if well marked and long, offers certain possibilities for a direct study of WU without the use of any formula, granting that one can demonstrate that ordinary learning is not taking place during the period of the plateau, or, in case it is taking place, one can point out that it does not show itself in the WU values. There is certainly considerable evidence that learning is going

on during the plateau period *in some cases.* But, granting that plateaus are periods of learning, one may well argue that since the effects of learning are not measurable, these effects need not be considered in the study of WU. Again one may note that it has never been proved that all or even the majority of plateaus found in learning curves are actually periods of learning.

If one can assume that learning-effects are negligible in the case of well marked plateaus one can proceed to compare directly the records of various stages of the day's work. In our work on WU (using the formboard) we have calculated the value of WU in the case of records obtained during periods of well marked and relatively flat plateaus, by comparing directly the first and last halves of each day's work. We have taken these same records and calculated WU by the use of the Robinson-Heron formula. The results of the methods are shown in Table I.

TABLE I

Comparative Values of WU when Computed Directly and by Means of the Robinson-Heron Formula. All Records Used Occurred During Periods of Plateaus

Subject	Direct Comparison	Using R-H Formula
S	Mean 0.32	Mean 0.27
	Range 0.04-0.88	Range 0.04-0.56
	71% show WU	59% show WU
	(24 Comparisons)	(22 Comparisons)
P	Mean 0.27	Mean 0.22
	Range 0.06-0.78	Range 0.06-0.44
	100% show WU	75% show WU
	(19 Comparisons)	(16 Comparisons)

A study of Table I listed above indicates that, on the whole, more WU is found when the direct method of comparison is made between early and later records in a day's work than when the Robinson-Heron formula is used. However, the validity of this method of direct utilization of well-marked plateaus remains questionable.

Other Problems of Measuring WU

A very complexing problem which any method of treatment must encounter is that concerning the nature and definition of learning itself. In studying subjects' protocols and their objective records the writer has been forced to speak of "form." We have used the term for want of a better word and in the sense in which it is usually used in sports. The highly trained expert in most any sport may find himself momentarily or for longer periods of time "off form." In working the formboard our well trained subjects at times indicated this "off form" at the very beginning of the work. They would then hit upon or fall into, so to speak, their correct form. Shall we call this short, often only momentary, period of "off form" a period of learning? Is the process of regaining proper form to be called learning?

In his excellent little book on *Conditions and Consequences of Human Variability,* Dodge[6] reports something of the same phenomenon, a period of initial adjustment, in the case of rapid finger oscillation and in the case of the pursuit movement of the eyes as they followed an oscillating object. Speaking of the pursuit of movement, Dodge writes, "Like the oscillation of a finger, even such a simple and practiced act as the ocular pursuit of an object in harmonic oscillation does not start perfectly formed in successive instances, but shows a temporal development of approximation and correction in each case."

Dodge regards this brief initial period of adjustment as a case of *learning,* at least as an elementary learning process. It is difficult to know just how much should be included under learning. One might regard such initial periods of adjustment as analagous to a blocked recall in the case of images and ideas. No new functional con-

[6] Raymond Dodge. Conditions and Consequences of Human Variability, *Yale University Press,* 1931. Quotation from page 88.

nections are set up. Rather, one might regard such periods as periods during which a proper total stimulating situation is developing, involving perhaps many receptors, and during which an existing inhibiting neural "set-up" fades away.

If one agrees with Dodge in calling these initial periods of adjustment learning, then one must say that at every successive act on the part of any expert an initial period of *re-learning* takes place. In which case any initial retardation in work-efficiency in any given act would be a case of learning and if we use the term WU it would mean just this fact of initial re-learning.

A second important question in any technique of measuring and analyzing WU concerns the size of the work-increments which are to be measured and compared. We shall refer once more to the formboard which we have used in most of our own work on WU. Let us suppose that the subject works the board twenty times. Should one compare the first and last trials of the day's work, the first and last two trials, the first and last five trials, the first and last ten trials, or yet some other arrangement? The writer has found that if one compares the first and last trials, or the first and last two trials, considerably more WU is shown than when one compares the first and last half of the day's work. We have also found that WU may take place *within the first trial* itself. Time and again it was observed that the subject was working more efficiently toward the end of the trial than at the beginning. The introspective reports substantiated this observation.

It is possible, then, that no WU may be shown in any task of a more complicated nature for the very simple reason that the unit of work or measurement is too large. In general it seems to be a safe principle to measure very small units of work in their successive order of appearance. Conversely, the larger the increment of work which is used as a unit the more may WU

effects be hidden within the unit of work. While, from a practical point of view, extremely small amounts of WU may be of no importance, from a theoretical point of view, the adequate discovery and measurement of these small amounts of WU are of great importance.

AN ECONOMICAL METHOD OF ADMINISTERING AND SCORING NEW-TYPE EXAMINATIONS

PAUL J. FAY and WARREN C. MIDDLETON
DePauw University

The use of objective tests and examinations constitutes a serious problem for instructors in institutions with limited mimeographing service. The problem becomes especially acute in general and educational psychology courses with large enrollments—courses in which the new-type test is often employed as the sole basis of measurement. The usual method of administering objective examinations involves such expense that their extensive use is almost prohibited in institutions with small departmental budgets. This expense undoubtedly has caused many psychology instructors in recent years to substitute dictated tests for mimeographed tests.[1]

The high cost of mimeographed tests and examinations is due chiefly to the fact that a test on which a student writes or marks his responses cannot be used a second time; the test sheet is destroyed in the taking. The scoring of mimeographed tests containing several pages is likewise time-consuming. In order to decrease the expense in money and time of mimeographed examinations, the writers have been experimenting for the last two years with a simple testing method—a method which has proved satisfactory from several standpoints. The procedure has been tried in several elementary and advanced psychology courses. Its greatest application

[1] Although there are a few studies which indicate that dictated tests are about as reliable as mimeographed or printed tests, it would still seem to be desirable—for several reasons—to use visual rather than auditory methods.

was naturally found in such courses as general, educational, and abnormal psychology—that is, in courses which have large enrollments and which require frequent testing.

I. Administering the Test

The first step, of course, in administering a test is to have the examination questions mimeographed. It is wise to provide a generous supply of the original examination papers; papers are easily soiled or mutilated by frequent handling, and duplicates of the original must be occasionally added.[2]

Have mimeographed on eight and one-half by eleven inch paper a large supply[3] of blank answer-sheets similar to Figure I. These sheets should be so made that they will accommodate answers to 200 test items on each side; this makes it possible to use them for any length examination. Answers are not written or marked on the question sheets, but on this especially devised answer-sheet. Consequently, the original examination can be used almost indefinitely—at least until a change in class lectures or textbook necessitates a new examination.

Answers are recorded on the answer-sheet merely by using certain signs or letters. For true-false questions + and 0 signs are used, and for all other forms of new-type tests capital letters (printed) of the alphabet are employed. In the case of a matching test or a modified form of the completion test the writers have found that test groups of twenty-five items are most satisfactory; these allow the use of letters from A to Y. If the in-

[2] The writers have found that students ordinarily will not soil or mutilate test papers, if they are repeatedly cautioned not to do so. It may be advisable to penalize students who carelessly mutilate papers.

[3] It is perhaps wise to mimeograph three or four thousand answer-sheets at one time, since they disappear quickly when used in an intensive testing program. By omitting a course number, the sheets can be used for any course.

FIGURE I

Front of Answer Sheet

Name_____

Psychology_____ No._____

1.	26.	51.	76.	101.	126.	151.	176.
2.	27.	52.	77.	102.	127.	152.	177.
3.	28.	53.	78.	103.	128.	153.	178.
4.	29.	54.	79.	104.	129.	154.	179.
5.	30.	55.	80.	105.	130.	155.	180.
6.	31.	56.	81.	106.	131.	156.	181.
7.	32.	57.	82.	107.	132.	157.	182.
8.	33.	58.	83.	108.	133.	158.	183.
9.	34.	59.	84.	109.	134.	159.	184.
10.	35.	60.	85.	110.	135.	160.	185.
11.	36.	61.	86.	111.	136.	161.	186.
12.	37.	62.	87.	112.	137	162.	187.
13.	38.	63.	88.	113.	138.	163.	188.
14.	39.	64.	89.	114.	139.	164.	189.
15.	40.	65.	90.	115.	140.	165.	190.
16.	41.	66.	91.	116.	141.	166.	191.
17.	42.	67.	92.	117.	142.	167.	192.
18.	43.	68.	93.	118.	143.	168.	193.
19.	44.	69.	94.	119.	144.	169.	194.
20.	45.	70.	95.	120.	145.	170.	195.
21.	46.	71.	96.	121.	146.	171.	196.
22.	47.	72.	97.	122.	147.	172.	197.
23.	48.	73.	98.	123.	148.	173.	198.
24.	49.	74.	99.	124.	149.	174.	199.
25.	50.	75.	100.	125.	150.	175.	200.

structor wishes to use more than twenty-five items in a single test group, he may designate the additional answers by doubled letters, such as AA, MM, RR.

Practically any of the usual types of objective examinations can be utilized with this method of testing. The writers have found that the following types can be satisfactorily employed:

(1) *True-False.* In the true-false test, answers to the questions 1, 2, 3, 4, etc., are marked respectively on the answer-sheet either + (right) or 0 (wrong).[4]

(2) *Matching.* In the matching test, the items to be matched are arranged on the examination sheet in parallel columns, the numbered items being placed on the left side of the sheet and the lettered items on the right side of the sheet. The mimeographed directions given to the student are in some such form as the following:

After the number on your test sheet corresponding to the number of each item in Column I, write the *letter* of the item in Column II that is most closely related to it. Do not use any letter more than once.

(3) *Subject-Predicate.* The subject-predicate test is a special form of the matching test and is answered in the same manner. Instead of having the two parts to be matched placed in vertical columns, however, it is better to place the numbered "subject" items at the top of the sheet and the lettered "predicate" items on the lower part of the sheet. The directions given to the student are in some such form as the following:

After the number on your test sheet corresponding to that of each subject in Part I, write the *letter* of the predicate in Part II

[4] The writers have found + and 0 more satisfactory than the letters T and F, since the latter too often resemble each other when not printed and capitalized. Also, the + and 0 marks are superior to + and —, since the vertical line of the + mark is often indistinct and confusion in scoring naturally results.

that best completes the sentence. Do not use any letter more than once.

(4) *Mutilated Sentences.* The mutilated-sentences test (a modified form of the completion test) can be used by mimeographing on the examination sheet all the words to be inserted in the incomplete sentences. The words are placed at the top of the sheet and are lettered in mixed order. The mutilated sentences are placed below and the blanks are numbered consecutively. The directions given to the student are in some such form as the following:

The words listed under Group I belong in the blanks left in the sentences under Group II. After the number on your test sheet corresponding to that of each blank in Group II, write the *letter* of the word in Group I that *most correctly* completes the sentence. Do not use any letter more than once.

(5) *Cross-Out.* A modified form of the cross-out test lends itself readily to this form of testing. Each item consists of five parts, lettered, respectively, (A), (B), (C), (D), and (E). The directions given to the student are in some such form as the following:

In each of the following groups of names, words, or phrases one name, word, or phrase does not belong in the series with the other four. After the number on your test sheet corresponding to the number of each group on this sheet, write the *letter* of the name, word, or phrase not belonging (or *least* belonging) with the other four of the series.

(6) *Best-Answer.* In the best-answer test each item consists of an incomplete statement followed by five (or more) alternative parts lettered (A), (B), (C), etc. The directions given to the student are in some such form as the following:

Each of the following incomplete sentences is followed by five alternative endings: (A), (B), (C), etc. After the number on your test sheet corresponding to the number of each sentence,

write the *letter* of the ending which makes the best and truest sentence.

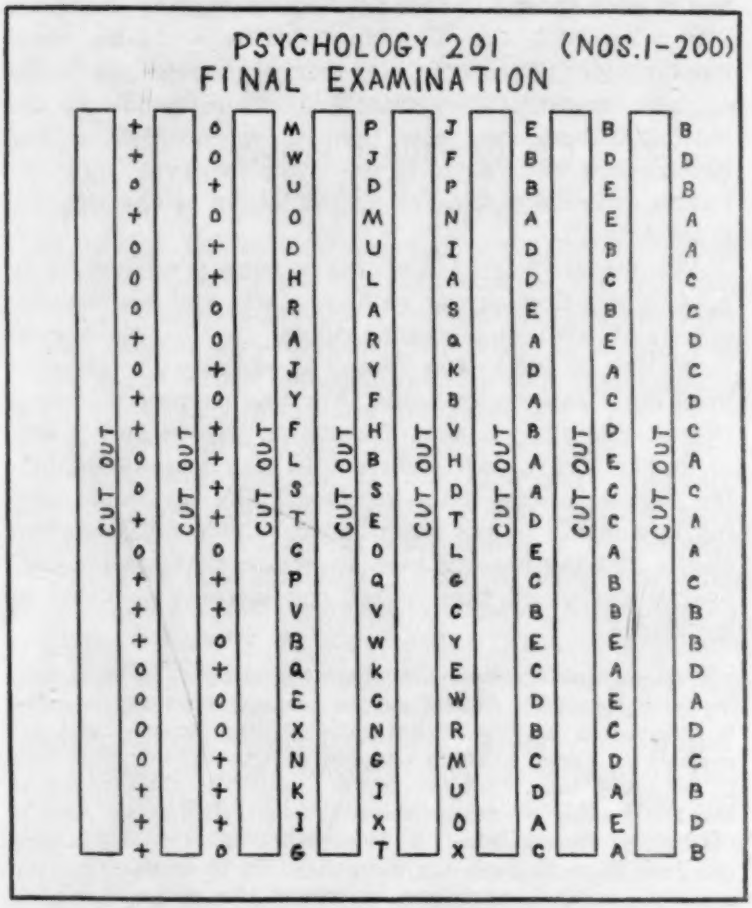
II. Scoring the Test

Speed and ease in scoring are two desirable features of a test that are always appreciated by examiners. Economy of time and economy of effort are to be sought in any testing procedure; otherwise, an instructor tends to become merely a clerical worker. The testing technique here described makes possible a quick, easy, and accurate scoring of objective examinations. The writers have achieved the most satisfactory results with a method similar in its essential features to that used in the Strong Vocational Interest Test. A scoring key for a 200-item test is illustrated in Figure II. It is constructed from an eight and one-half by eleven inch cardboard; thus, it is the same size as the answer-sheet illustrated in Figure I. Eight grooves, each one-half inch wide and nine and one-quarter inches long, are cut in the cardboard. These grooves are so spaced that they fall over the blank spaces between the numbers on the answer-sheet. Enough blank keys can be constructed at the beginning of the year for all the tests and examinations of the year. The answer letters, or signs, are written in India ink[5] to the right of the grooves, so that each key sign is placed exactly opposite its corresponding answer on the answer-sheet beneath. By correctly answering one test sheet and placing a blank scoring key over it, the signs can readily be written on the key. A test of any number of items from 1 to 200 can be scored with one side of the cardboard scoring key; by reversing the cardboard, items from 201 to 400 can be scored.

[5] India ink should be used, since it does not easily fade out or smear. The letters can be made very legible by using a regular lettering pen.

FIGURE II.

Front of Scoring Key for Final Examination



The ease and speed of scoring permitted by this method will be readily apparent, since 200 test items can be scored without raising the scoring key from the answer-sheet. The writers have found that, using a 200 mixed-item examination, thirty to forty answer-sheets can be scored per hour. This time includes checking

errors, counting the number of errors, mentally subtracting this number from 200, and writing the score at the top of each sheet. It does not include copying the scores into a record book. The true-false tests can be scored most rapidly, since only two marks are used. It takes slightly longer to score matching, subject-predicate and mutilated sentences tests than to score cross-out and best-answer tests, since in the latter two types only five letters are used instead of all the letters of the alphabet from A to Y.

The writers have found this scoring procedure to be particularly time-saving with quarterly and midsemester tests and with final examinations—that is, with tests consisting of 100 to 400 items. It can also be used with brief daily and weekly tests. For this purpose, however, the writers suggest that the answer sheets be cut into narrower strips, just sufficient in size to accommodate the number of test items involved. The cardboard scoring key should be cut accordingly. Cutting the answer-sheets and the scoring keys to fit the examination will, over a period of time, effect considerable economy of material.[6]

[6] Mimeographed answer-sheets are not essential in short tests involving from 25 to 100 items. An economical testing procedure is achieved by cutting regular ruled writing paper (eight and one-half by eleven inches in size and with twenty-five ruled lines to the page) into strips two inches wide. The students can be asked to number the answer-sheets themselves before the start of the test, numbering from 1 to 25 down the left edge of the strip and from 26 to 50 down the right edge. If 75 or 100 test items are to be used, the strip can be turned over and numbered in the same way from 51 to 100. The papers can be quickly and accurately scored with a cardboard scoring key.

THE BELIEFS OF COLLEGE STUDENTS CONCERNING EVOLUTION

GEORGE J. DUDYCHA
Ripon College

Since all college students are sooner or later confronted with ideas concerning evolution and development, what are their attitudes toward these ideas—what are their beliefs concerning evolution? In order to answer this question the writer obtained answers to a questionnaire from 1054 college Freshmen and Seniors.

The Freshmen who matriculated at a number of mid-western colleges[1] in the fall of 1930 were subjected during the registration period or during the first week of instruction to a series of tests, one of which included beliefs concerning evolution. Since the Freshmen made their responses to the propositions before they attended a single lecture or during the first week of academic instruction, their judgments were not influenced by the college but were made in accord with the beliefs which they brought to college from their various communities. The Seniors of seven colleges[2] who participated in this investigation were confronted in the spring of 1930, and thus almost at the close of their college course, with the same set of propositions. Since the tests were administered near the close of the students' senior year their judgments were influenced by four years of college instruction.

The procedure used in this study was as follows. Each

[1] Baldwin-Wallace, Coe, Milton, Monmouth, Ripon and Dubuque University.

[2] Baldwin-Wallace, Coe, Huron, Illinois, Milton, Monmouth and Ripon Colleges.

student was confronted with a printed list of twenty-five propositions concerning evolution. Each proposition was stated briefly, clearly and affirmatively. To the right of the statement of each proposition were five groups of parentheses labeled from left to right, A, B, C, D and E. The list of propositions was prefaced with the following instructions:

> Opposite each of the following propositions make a cross (X), in one of the five columns, according to your belief.
> Under A if you implicitly believe.
> Under B if you are inclined to believe, but doubt.
> Under C if you do not know whether you believe or do not believe (non-committal).
> Under D if you are inclined to disbelieve and doubt.
> Under E if you absolutely do not believe.

Thus any one of five possible answers could be given to each proposition. This is a distinct advantage for the student could not only express indifference or a non-committal attitude and belief or disbelief but also within certain limits the extent of his belief or disbelief. The advantage over the yes-no type of statement is obvious. For a detailed description of the method used for determining the validity of the results, see one of the writer's earlier reports.[3]

First let us consider the results obtained for college Freshmen. If the reader will direct his attention to Table I, he will immediately observe the one outstanding characteristic of these results—the rather large per cent of non-committal responses. Each proposition was marked non-committal by at least 10 per cent of the Freshmen and one by as high as 43 per cent, with an average of 24 per cent. Thus, on the average, practically one-fourth of the responses were non-committal responses. This high percentage may be taken to indicate one of two things: either the Freshmen tested were

[3] "The Religious Beliefs of College Freshmen," *School and Society*, 31:206-208, February 8, 1930.

reluctant about expressing their beliefs concerning evolution, or else they lacked the necessary information and thus were not able to express intelligent judgments of belief or disbelief concerning some of the propositions presented. Yet another item of interest may be observed. The range of belief is from 71 down to 6 per cent and the range of disbelief is from 6 to 69 per cent. Although each statement was believed or disbelieved by at least 6 per cent of the 750 freshmen, no statement was believed or disbelieved by practically all of the students.

The statement which ranks first, believed by 71 per cent of the Freshmen, is, "Evolution is a description of nature's process of development;" the one which ranks second, believed by 67 per cent, is, "The universe is still evolving," and the one in third place, believed by 65 per cent, is, "Evolution accounts for the development of life from lower to higher forms, or from simple to complex." Although a fairly large percentage of the Freshmen believed these first three propositions, each of which favors belief in evolution, they are rather evenly divided between belief and disbelief in evolution itself, as is evident from their response to proposition twelve, "Evolution" (believed by 35 per cent and disbelieved by 36 per cent.)

Some other propositions which were believed by more than 50 per cent of the students are: "Stars, societies and institutions have evolved as well as animals" (57 and 12 per cent); "Evolution is a process of creation" (58 and 20 per cent); "Evolution is a necessary idea in science" (55 and 24 per cent). One proposition which we might expect to rank high but which was placed low in the list, is the nineteenth, "There isn't any doubt about the fact of evolution" (21 and 47 per cent). Scientists agree that there isn't any doubt about the fact of evolution, but that there is a question as to the manner and means of evolution. This distinction was apparently not made by the Freshmen.

TABLE I

College Freshmen

List of Propositions Concerning Evolution and Distribution of Responses Indicated by Per Cent (750 cases)

	Propositions	A	B	C	D	E
1.	Evolution is a description of nature's process of development	59	12	20	02	07
2.	The universe is still evolving	57	10	27	02	04
3.	Evolution accounts for the development of life from lower to higher forms, or from simple to complex	53	12	26	02	07
4.	Evolution teaches that man is related to the lower animals	53	13	17	05	12
5.	A knowledge of evolution broadens one's outlook on life	51	13	21	04	11
6.	Evolution is accepted more extensively today than 25 years ago	51	11	22	04	12
7.	Stars, societies and institutions have evolved as well as animals	44	13	31	02	10
8.	Evolution is a process of creation	46	12	22	03	17
9.	Evolution is a necessary idea in science	41	14	21	06	18
10.	Chas. Darwin was the originator of the evolution idea	43	10	20	04	23
11.	Evolution teaches that man came from monkey	36	10	10	04	40
12.	Evolution	23	12	29	08	28
13.	Evolution is diametrically opposed to all religious teachings	26	10	26	07	31
14.	No true Christian believes in evolution	24	08	21	08	39
15.	Evolution pertains only to biology and specie formation	19	10	31	06	34
16.	God has no place in any of the theories of evolution	21	07	27	09	36
17.	Evolution should be taught in every American college	17	09	32	09	33
18.	Evolution is true only in so far as it refers to the growth of an organism as from egg to hen	14	05	43	07	31
19.	There isn't any doubt about the fact of evolution	14	07	32	09	38
20.	Evolution is merely a passing fancy in science	11	09	31	10	39
21.	One can gain nothing from a study of the theories of evolution	12	06	16	10	56
22.	All scientists are evolutionists	08	07	19	09	57
23.	One is certain to lose his religion if he studies evolution	06	07	18	13	56
24.	Darwin's theory is the only theory of evolution	03	03	31	09	54
25.	Evolution is a teaching of the devil	05	04	22	07	62

Obviously the beliefs of the Freshmen concerning evolution are significant only to the extent to which the Freshmen show insight into the meaning, implications and known facts concerning this concept. Let us see if their information tends to be correct. Each of the following statements, by its comparatively large percentage of belief or disbelief, indicates that the beliefs of the Freshmen are quite in accord with those of present-day scientists and philosophers. These propositions are: "Evolution teaches that man is related to the lower animals" (66 and 17 per cent); "Evolution is accepted more extensively today than 25 years ago" (62 and 16 per cent); "Stars, societies and institutions have evolved as well as animals" (57 and 12 per cent); "Evolution is a process of creation" (58 and 20 per cent); "Evolution pertains only to biology and specie formation" (29 and 40 per cent); "God has no place in any of the theories of evolution" (28 and 45 per cent); "Evolution is true only in so far as it refers to the growth of an organism as from egg to hen" (19 and 38 per cent); "All scientists are evolutionists" (15 and 66 per cent), and "Darwin's theory is the only theory of evolution" (6 and 63 per cent). There are two statements, however, to which we must call attention because the belief of the Freshmen in these two propositions indicates somewhat faulty information. These are the tenth—"Charles Darwin was the originator of the evolution idea" (53 and 27 per cent), and the eleventh—"Evolution teaches that man came from monkey" (46 and 44 per cent). Although Charles Darwin is popularly referred to as the originator of the evolution idea, he merely reformulated and attempted to substantiate an idea which we today know dates back to early times. As for the idea that evolution teaches that man came from monkey, we can merely say that this is an idea which has unfortunately become ingrained in the thinking of the layman.

Another point of view from which we desire to examine our results is from the angle of the open-mindedness of the Freshmen. Are the Freshmen examined open-minded or are they prejudiced against receiving any instruction concerning evolution? The propositions which will reveal this are: the fifth, "A knowledge of evolution broadens one's outlook on life" (64 and 15 per cent); the twentieth, "Evolution is merely a passing fancy in science" (20 and 49 per cent); the twenty-first, "One can gain nothing from a study of the theories of evolution" (18 and 66 per cent), and the seventeenth, "Evolution should be taught in every American college" (42 and 26 per cent). Belief in the first three of these statements certainly indicates an open-minded attitude. We may draw two conclusions, however, from the attitude of the Freshmen toward the fourth proposition. Either the Freshmen indicate a narrow-minded attitude, which is contrary to their belief in the three statements just discussed, or else a conservative attitude which is in accord with their general conservatism. Perhaps they feel that in view of the fact that evolution is still quite disturbing to some it should not be taught in all American colleges. Which of these conclusions is the correct one we can not say.

There is yet another group of statements which we desire to examine, the statements dealing with evolution and religion. These statements are the thirteenth, "Evolution is diametrically opposed to all religious teachings" (36 and 38 per cent); the fourteenth, "No true Christian believes in evolution" (32 and 47 per cent); the sixteenth, "God has no place in any of the theories of evolution" (28 and 45 per cent); the twenty-third, "One is certain to lose his religion if he studies evolution" (13 and 69 per cent), and the twenty-fifth, "Evolution is a teaching of the devil" (9 and 68 per cent). As in the case of "Evolution," we note that the Freshmen are rather evenly divided between belief and disbelief with regard

to the ideas that evolution is diametrically opposed to all religious teachings. In spite of this division of the Freshmen they were inclined to disbelieve that no true Christian believes in evolution. Disbelief in the twenty-third proposition indicates a broad-minded and wholesome attitude toward both religion and evolution.

Let us summarize briefly. They place the propositions favoring belief in evolution at the head of the list and indicated rather extensive belief in them, even though they are rather evenly divided with regard to belief in evolution itself. Their belief and disbelief in certain propositions indicate that for the most part their information concerning the principles of evolution is correct; only in two cases does their belief indicate faulty information and in both widely-accepted popular beliefs are involved. Their attitude toward evolution and its bearing on religion is one of open-mindedness. They do not seem to experience any pronounced conflict between their ideas concerning evolution and their religious beliefs. On the whole we may say that their attitude toward evolution is favorable rather than unfavorable.

Now let us turn our attention to college Seniors and see what they believe concerning evolution. As we glance at the results found in Table II we immediately note that there is a wide range of belief and disbelief, and also that whereas some of the statements are believed by a large percentage of the Seniors, others are disbelieved as extensively. Belief extends from 91 down to 1 per cent; disbelief extends from 1 to 93 per cent.

We find, further, that belief in the first seven propositions is above 80 per cent. Belief in the first proposition, "The universe is still evolving" (91 and 1 per cent); the second, "Evolution is a description of nature's process of development" (86 and 5 per cent); the fourth, "Evolution accounts for the development of life from lower to higher, or from simple to complex" (84 and 5 per cent); the fifth, "Stars, societies and institutions

TABLE II

College Seniors

List of Propositions Concerning Evolution and Distribution of Responses Indicated by Per Cent (304 cases)

	Propositions	A	B	C	D	E
1.	The universe is still evolving	82	09	08	00	01
2.	Evolution is a description of nature's process of development	78	08	09	01	04
3.	A knowledge of evolution broadens one's outlook on life	75	10	09	03	03
4.	Evolution accounts for the development of life from lower to higher forms, or from simple to complex	76	08	11	01	04
5.	Stars, societies and institutions have evolved as well as animals	73	08	12	02	05
6.	Evolution is accepted more extensively today than 25 years ago	72	10	08	03	07
7.	Evolution is a necessary idea in science	65	16	08	03	08
8.	Evolution	61	14	18	02	05
9.	Evolution teaches that man is related to the lower animals	55	14	15	05	11
10.	Evolution should be taught in every American college	53	09	24	06	08
11.	Evolution is a process of creation	54	08	13	04	21
12.	There isn't any doubt about the fact of evolution	34	16	18	07	25
13.	Charles Darwin was the originator of the evolution idea	28	10	15	08	39
14.	All scientists are evolutionists	15	09	19	08	49
15.	Evolution pertains only to biology and specie formation	13	04	15	11	57
16.	Evolution is true only in so far as it refers to the growth of an organism as from egg to hen	04	03	21	08	64
17.	Evolution is diametrically opposed to all religious teachings	06	04	10	11	69
18.	Evolution teaches that man came from monkey	07	04	06	07	76
19.	Darwin's theory is the only theory of evolution	02	01	16	08	73
20.	God has no place in any of the theories of evolution	05	02	08	07	78
21.	Evolution is merely a passing fancy in science	02	03	11	11	73
22.	One is certain to lose his religion if he studies evolution	04	02	08	07	79
23.	No true Christian believes in evolution	02	04	07	07	80
24.	One can gain nothing from a study of the theories of evolution	04	02	04	08	82
25.	Evolution is a teaching of the devil	00	01	06	03	90

have evolved as well as animals" (81 and 7 per cent), and the seventh, "Evolution is a necessary idea in science" (81 and 11 per cent) favors belief in evolution. In spite of this fact, proposition eight, "Evolution," was believed by only 75 per cent of the Seniors. Apparently many of the Seniors are willing to express belief in the teachings of evolution, but some at least are much less willing to express belief in evolution itself. Another proposition which was ranked lower than we might expect in view of the extensive belief expressed in the first seven is twelve, "There isn't any doubt about the fact of evolution" (50 and 32 per cent). Apparently some of the students did not appreciate the distinction which is here involved.

The Seniors' extensive belief or disbelief in a number of propositions which follow indicates that they have insight into the implications and teachings of evolution and that their information tends to be correct. The propositions are: the fifth, "Stars, societies and institutions have evolved as well as animals" (81 and 7 per cent); the sixth, "Evolution is accepted more extensively today than twenty-five years ago" (82 and 10 per cent); the ninth, "Evolution teaches that man is related to the lower animals" (69 and 16 per cent); the eleventh, "Evolution is a process of creation" (62 and 25 per cent); the thirteenth, "Charles Darwin was the originator of the evolution idea" (38 and 47 per cent); the fifteenth, "Evolution pertains only to biology and specie formation" (17 and 68 per cent); the sixteenth, "Evolution is true only in so far as it refers to the growth of an organism as from egg to hen" (7 and 72 per cent), and the nineteenth, "Darwin's theory is the only theory of evolution" (3 and 81 per cent). Only in the case of the thirteenth proposition is there but a small difference between belief and disbelief, and this is probably due to a lack of stress upon the historical development of the idea of evolution.

Another phase of the attitude of college Seniors in which we are interested is with regard to their open-mindedness. The propositions which indicate this are: the third, "A knowledge of evolution broadens one's outlook on life" (85 and 6 per cent); the tenth, "Evolution should be taught in every American college" (62 and 14 per cent); the twenty-first, "Evolution is merely a passing fancy in science" (5 and 84 per cent), and the twenty-fourth, "One can gain nothing from a study of the theories of evolution" (6 and 90 per cent). In the case of each proposition, except the tenth, the percentage of belief and disbelief is such as to indicate that many of the Seniors who were unwilling to commit themselves to belief in evolution were, nevertheless, open-minded with regard to it.

The last group of propositions to which we must direct our attention is the one including statements which deal with evolution and religion. These statements are: the seventeenth, "Evolution is diametrically opposed to all religious teachings" (10 and 80 per cent); the twentieth, "God has no place in any of the theories of evolution" (7 and 85 per cent); the twenty-second, "One is certain to lose his religion if he studies evolution" (6 and 86 per cent); the twenty-third, "No true Christian believes in evolution" (6 and 87 per cent), and the twenty-fifth, "Evolution is a teaching of the devil" (1 and 93 per cent). The significant thing is that each of these propositions which states that religious beliefs and ideas concerning evolution are incompatible are disbelieved by 80 or more per cent of the 304 Seniors who indicated their beliefs. We may conclude from this that these particular Seniors have apparently succeeded quite well in harmonizing their religious and evolutionary ideas.

In brief, we have found that a large percentage of the 304 Seniors ranked the propositions favoring belief in evolution at the head of the list. Although the percentage of the group that believed these propositions

was great, only three-fourths of the Seniors were free to indicate their willingness to believe in evolution. It is evident from their responses to a number of the propositions that their information concerning evolution is correct and that their attitude is one of open-mindedness. Lastly we have observed that for these particular Seniors religion and evolution are not incompatible.

There are some interesting likenesses and differences between the beliefs of Freshmen and Seniors with regard to evolution. In the first place there is a great difference between the extent of belief and disbelief of the two groups. Whereas no proposition was believed by more than 71 per cent of the Freshmen, six statements were believed by more than 80 per cent and one by 91 per cent of the Seniors. On the other hand, although no proposition was disbelieved by more than 69 per cent of the Freshmen, seven propositions were disbelieved by more than 80 per cent and two by more than 90 per cent of the Seniors. This, of course, gives rise to a correspondingly great percentage of non-committal responses for the Freshmen than for the Seniors. We may conclude from these results that either the Seniors possessed such information concerning evolution that they could confidently express belief or disbelief in the particular propositions, which information the Freshmen lack, or that due to training the Seniors have outgrown their prejudices and hence expressed belief or disbelief frankly, whereas the Freshmen have not become emancipated from the prejudices and hence hesitated to commit themselves.

In spite of the differences which we have just noted there are a number of propositions which were given the same rank, and others which were given practically the same rank by both groups. The propositions which were given the same rank are the sixth, fifteenth and twenty-fifth. Also we note that both groups ranked, "The universe is still evolving" and "Evolution is a

description of nature's process of development" high, and "One can gain nothing from a study of the theories of evolution," "One is certain to lose his religion if he studies evolution" and "Evolution is a teaching of the devil" low.

Although we have noted that the Freshmen were *inclined* to believe, and that the Seniors were *greatly inclined* to believe in a number of the propositions which favor belief in evolution, both Freshmen and Seniors were much less willing to express belief in evolution itself.

We have observed earlier that the responses of many of the Seniors and of over half of the Freshmen to certain of the propositions indicate that their information about the nature and implications of evolution tend to be correct. The Seniors, however, by the extent of their belief indicate greater certainty with regard to the correctness of their information. Also we note from the responses of the members of the two groups that, although both are open-minded, the Seniors are more so.

Some of the propositions concerning which we find a marked divergence between the beliefs of the Freshmen and the Seniors are: "Evolution is diametrically opposed to all religious teachings," "No true Christian believes in evolution," and "God has no place in any of the theories of evolution."

On the whole we must conclude that the Seniors indicate by the nature of their responses that they are better informed, more open-minded and more inclined to believe in evolution than the Freshmen and that they apparently experience less of a conflict between their religious ideas and evolution than do the Freshmen.

FAMILIARITY WITH THE NAMES OF MUSICIANS[1]

PAUL R. FARNSWORTH
Stanford University

It may safely be said that psychologists can measure but poorly the effects of the environment. To compensate for this lack they are forced to work with large groups of subjects with the hope that environmental differences will counterbalance and that the similarities will be operative. With the exception of a few questionnaire studies, almost no attempts have been made to demonstrate the environmental similarities in the field of music. It was the aim of this study, therefore, to make some slight beginning in an analysis of the amount of familiarity shown for the names of musical personages.[2,3]

A list of 90 names culled from musical sources was given to 5 groups of subjects. It was estimated that approximately 30 of these names would probably be

[1] The writer is indebted to Miss Jane Loomis and to Mr. Minchi Young for aid in this work.

[2] It should be noted that this questionnaire does not *necessarily* measure "true" acquaintance with the musicians' names.

[3] Some time ago the writer found considerable agreement between large groups of subjects in their preferences for the intervals and phrases found in the Seashore Consonance measure (*r*, .96) and the Kwalwasser Melodic and Harmonic Sensitivity tests (*r*'s, .98 and .99). See *Genet. Psychol. Monog.*, 1931, *9*, 5, pp. 352, 356, and 357. This earlier work does not present as conclusive an argument for the uniformity of environmental effects as do the findings of the present paper for it has been maintained (but probably cannot be proved) that preferences may in some slight degree be determined by original nature.

rated as well known, approximately 30 as fairly well known, and 30 as relatively unknown. The subjects were chosen at random from each of the four years of high school and from sophomore classes in elementary psychology at Stanford. There were approximately 100 at each of the high school levels and 200 from college. The high school from which the subjects were drawn is located at Palo Alto, California. Work in music in this school is not compulsory but quite commonly elected. At Stanford there are no formal music courses. Approximately 5 per cent of the freshmen entering this university come from the Palo Alto High School.

The questionnaire is given below. The ranks, as found by tabulating the single checks, are here given with the questionnaire to conserve space.

QUESTIONNAIRE AND TABLE I

This questionnaire is being used to ascertain what musicians are known. Check only the names which are known to you. Put 2 checks in front of the four best known.[1]

Fr	So	Ju	Se	C	
66	81	80.5	47	60	Abt, F.
66	52.5	61.5	43	50.5	Ackermann, A. J.
82.5	52.5	61.5	47	42	Alberti, D.
82.5	81	80.5	84	82.5	Allihn, H.
35	65	46.5	43	43	Amati, V.
82.5	65	46.5	84	76	Amon, B.
66	81	61.5	69	82.5	Ancot, J.
82.5	81	80.5	69	76	Apell, D.
82.5	81	80.5	69	88.5	Asola, G. M.
46	52.5	61.5	47	40.5	Auber, D.
66	81	80.5	84	88.5	Avison, C.
8	8.5	9.5	6	5	Bach, J. S.
54	81	61.5	69	67.5	Badarczewska, T.
46	65	39	69	53	Balfe, M. W.
46	42.5	61.5	69	71.5	Becker, G.
2	2	1	1.5	1	Beethoven, L.
29	30.5	25	31.5	32	Bohm, C.

[1] A rank of "1" is highest. The first four columns represent the 4 high school classes and the "C" column the college group.

Fr	So	Ju	Se	C	
82.5	81	80.5	84	67.5	Boisselot, X.
66	65	61.5	69	82.5	Bordese, L.
9	15.5	12	14.5	15	Brahms, J.
66	58	61.5	69	55.5	Büchner, E.
82.5	81	61.5	54.5	58	Buck, D.
54	81	80.5	69	64	Bull, J.
54	81	80.5	84	82.5	Cavos, C.
31	38	30	38.5	36	Charpentier, G.
3	5	7	3.5	6.5	Chopin, F.
32.5	38	46.5	38.5	37.5	Corelli, A.
82.5	65	80.5	54.5	67.5	Couperin, C.
46	81	35.5	33.5	45.5	Cramer, J. B.
66	81	80.5	84	44	Cui, C. A.
25	22	21	23	22	Debussy, C.
66	81	80.5	84	82.5	Delprat, C.
39	30.5	80.5	38.5	29.5	Donizetti, G.
54	81	39	69	58	Dorn, O.
82.5	81	61.5	69	88.5	Doss, A.
66	46	61.5	69	53	Dussek, J.
28	46	46.5	43	35	Elgar, E.
46	42.5	35.5	38.5	48.5	Field, J.
66	81	61.5	54.5	71.5	Fink, C.
66	52.5	46.5	84	61.5	Fiorillo, F.
82.5	65	46.5	43	40.5	Flotow, F. von
66	81	80.5	54.5	64	Gade, N. W.
35	28	30	26.5	27	Godard, B.
66	65	46.5	69	58	Goldmark, C.
46	34.5	33	35.5	45.5	Götze, O.
16.5	25	19	14.5	21	Gounod, C.
11.5	6.5	13	20	16	Grieg, E. H.
19	18	22.5	11.5	9	Handel, G. F.
39	32.5	46.5	54.5	61.5	Häser, H.
16.5	15.5	15.5	18	19	Haydn, F. J.
30	38	33	31.5	34	Heller, S.
39	34.5	39	69	55.5	Hummel, J. N.
32.5	38	33	30	37.5	Köhler, L.
82.5	52.5	80.5	84	82.5	Kücken, F. W.
46	52.5	80.5	54.5	39	Lange, G.
54	42.5	46.5	84	82.5	Langer, E.
19	12.5	15.5	14.5	14	Liszt, F.
13	19.5	22.5	22	23	MacDowell, E. A.
66	52.5	46.5	54.5	76	Mahler, G.
82.5	52.5	80.5	33.5	33	Mascagni, P.

Fr	So	Ju	Se	C	
26	21	28	29	25	Massenet, J.
11.5	3.5	8	1.5	6.5	Mendelssohn, F.
46	42.5	46.5	54.5	76	Metzdorff, R.
46	65	61.5	54.5	48.5	Monteverde, C.
27	27	27	28	26	Moszkowski, M.
4.5	3.5	4	3.5	8	Mozart, W. A.
21.5	19.5	18	18	13	Rachmaninoff, S.
82.5	65	80.5	84	64	Radecke, R.
39	32.5	61.5	43	53	Raff, J.
35	38	37	35.5	31	Ravel, M.
82.5	65	61.5	69	82.5	Riehl, W. H. von
14.5	14	15.5	14.5	20	Rimsky-Korsakov, N.
19	17	15.5	18	18	Rubinstein, A.
21.5	24	24	24	24	Saint-Saëns, C.
66	46	80.5	69	71.5	Schläger, H.
6	6.5	9.5	11.5	10	Schumann, R.
1	1	3	5	2	Schubert, F.
54	65	46.5	84	71.5	Schuett, E.
82.5	52.5	80.5	69	76	Serov, A. N.
4.5	10.5	2	10	3	Sousa, J. P.
66	65	80.5	54.5	47	Spohr, L.
14.5	12.5	11	9	12	Strauss, R.
23	29	26	26.5	28	Stravinsky, I.
82.5	65	61.5	84	67.5	Street, G. E.
10	10.5	6	7.5	11	Tschaikowsky, P. I.
66	52.5	61.5	54.5	50.5	Veracini, F. M.
24	23	20	21	17	Verdi, G.
7	8.5	5	7.5	4	Wagner, R.
54	26	30	25	29.5	Weber, C. M. von
39	81	61.5	54.5	88.5	White, M. V.

There seem to be no consistent progressive changes in the ranks except for Flotow and Rachmaninoff, and perhaps for MacDowell and Verdi.

The intercorrelations of the ranks are given in Table II. The raw scores obtained by the several musicians were employed in the correlations (product moment r's).

TABLE II[1,2]

	College	Senior	Junior	Sophomore	Freshman
College	.99	.99	.92	.94	.94
Senior			.96	.99	.97
Junior				.95	.97
Sophomore					.97

The 'senior,' 'junior,' 'sophomore' and 'freshman' refer to the high school groups. The first value to the left, .99, was obtained by splitting the college group into 2 sub-groups of 100 each, retabulating and intercorrelating the resultant raw scores.

It might be mentioned at this point that 8 names which appeared among the first 10 of the best known in the ranks of 1 of the sub-groups appeared among the first 10 of the other college ranking. (Handel and Schumann did not appear among the first 10 in the second sub-group's ranking. Their places were taken by Tschaikowsky and Rachmaninoff.)

An attempt was made to compare the '4 best known' or 'double check' method of tabulation with the 'single check' method. The college data were employed in this analysis. Twenty-eight of the 90 names were found to be double checked. The ranking of these correlated .87 (r obtained from rho) with the ranking obtained by the single check method. The 10 best known as found by this method of double checking were identical with those

[1] The P.E.'s are of the order .01 or less.

[2] The ranks and correlations depend to some extent on the musicians listed. If other names had been added, slight changes in the correlational values would most probably have occurred. Had the tabulations been different another source of change might have been introduced. It is conceivable that the large number of zero ballots (16 names did not receive a single check from the freshmen) might tend to render spurious to some extent the correlational values. To check this the correlations were rerun using only the names which had received at least 1 ballot from the lower academic group in each correlational pair. When this procedure was followed 3 of the coefficients were raised and the remainder were lowered. The maximum change, however, was of the magnitude .03 and the average change was less than .02. With such minute differences the general conclusions remain unchanged.

obtained by the single check method on 1 of the subgroups.

The following conclusions might safely be drawn from this study.

1. Two large unselected groups of college sophomores claimed the same order of acquaintance (r, .99) with the names of 90 musicians (deliberately chosen to include well known, quite commonly known and rarely known).

2. This uniformity of response was found to extend well down to the freshman high school level. (r's of .92 to .99.)

3. A double checking method to give the 4 best known names was found to render rankings quite similar (r, .87) to those obtained by the method of single checking in which all the known names were checked.

Presumably, then, the environment has produced considerable uniformity in this type of musical response.

ON ESTIMATING THE RELIABILITY COEFFICIENT

JOHN W. DICKEY
State Normal School, Newark, N. J.

T. L. Kelley[1] derives the formula for estimating the reliability coefficient of a test for one range, when the reliability coefficient of this test is known for another range. The formula is

$$\frac{\sigma}{\Sigma} = \frac{\sqrt{1-R}}{\sqrt{1-r}} \qquad (1)$$

where σ and Σ are the two known ranges, r and R the respective reliability coefficients one of which is known. Formula (1) when used is most easily expressed as

$$R = 1 - \frac{\sigma^2(1-r)}{\Sigma^2} \qquad (3)$$

or

$$r = 1 - \frac{\Sigma^2(1-R)}{\sigma^2} \qquad (5)$$

Dunlap and Cureton[2] derive the standard error formulas for these two expressions. They are for formulas (3) and (5)

$$\sigma_R = (1-R)\left[\frac{2}{N} + \frac{3-r^2}{n}\right]^{\frac{1}{2}} \qquad (7)$$

[1] Kelley, T. L., "The Reliability of Test Scores," *Journal of Educational Research*, Vol. III, No. 5, May, 1921, p. 337.

[2] Dunlap, Jack W., and Cureton, Edward E.: "Note on the Standard Error of a Reliability Coefficient Estimated from the Coefficient for a Different Range of Talent," *Journal of Educational Psychology*, Vol. XX, No. 9, Dec., 1929, p. 705.

and

$$\sigma_r = (1-r)\left[\frac{2}{n}+\frac{3-R^2}{N}\right]^{\frac{1}{2}} \qquad (9)$$

respectively.

The aim of this paper is to derive a new formula with its standard error formulas, which is to replace formula (1). This new formula proves to be superior to formula (1) because its standard error value proves to be smaller than that for formula (1).

We shall now derive the new formula (2). The writer[3] has derived the standard error of a "standard" score when taken as evidence of the best estimated "true" standard score to be

$$\sigma_{(z_1 - z_\infty)} = 1 - \sqrt{r}$$

where r is the reliability of the test. This derivation avoids a systematic error by using the best estimate of the "true" variability instead of the variability of the best estimated "true" scores. The standard error of a gross score when taken as evidence of the best estimated "true" gross score is then

$$\sigma_{(x_1 - x_\infty)} = \sigma_1(1 - \sqrt{r})$$

where σ_1 is the measure of one range of talent, and r is the corresponding reliability coefficient. Similarly

$$\Sigma_{(x_1 - x_\infty)} = \Sigma_1(1 - \sqrt{R})$$

where Σ_1 is the measure of another range of talent, and R is the corresponding reliability coefficient. Assuming

[3] Dickey, John W.: "On the Reliability of a Standard Score," *Journal of Educational Psychology*, Vol. XXI, No. 7, October, 1930, p. 547.

with Professor Kelley[4] that "the test is equally as effective in one range as in the other," that is, that

$$\Sigma_{1}{}_{(x_1 - x_\infty)} = \sigma_{(x_1 - x_\infty)}$$

we obtain directly

$$\sigma_1(1 - \sqrt{r}) = \Sigma_1(1 - \sqrt{R})$$

whence

$$\frac{\sigma}{\Sigma} = \frac{1 - \sqrt{R}}{1 - \sqrt{r}} \tag{2}$$

Here, as in formula (1), σ and Σ (neglecting subscripts) are the two known ranges, r and R are the respective reliability coefficients one of which is known. Formula (2) when used to evaluate R and r respectively, is most easily expressed as

$$R = \left[1 - \frac{\sigma(1 - \sqrt{r})}{\Sigma}\right]^2 \tag{4}$$

and

$$r = \left[1 - \frac{\Sigma(1 - \sqrt{R})}{\sigma}\right]^2 \tag{6}$$

To avoid imaginary numbers in actual substitution, we use the absolute values (disregard sign) of the given coefficients in formulas (4) and (6) and interpret the estimated coefficients accordingly.

The standard error formulas for expressions (4) and (6) will now be derived. These standard error formulas are essential; and for our immediate purpose they are necessary to establish the superiority of formula (2) over Kelley's formula (1).

[4] Holzinger, Karl J.: "On the Assumption that Errors of Estimate are Equal in Narrow and Wide Ranges," *Journal of Educational Research*, Vol. IV, No. 3, October, 1921, p. 237. Kelley's reply follows Holzinger's criticism.

Let,

n be the population of the shorter range, σ.

N be the population of the longer range, Σ.

Writing the total derivative of formula (4), we get

$$dR = \frac{2\sigma\sqrt{R}\,dr}{2\Sigma\sqrt{r}} + \frac{2\sqrt{R}(\sqrt{r}-1)\,d\sigma}{\Sigma} + \frac{2\sigma\sqrt{R}(1-\sqrt{r})\,d\Sigma}{\Sigma^2}$$

Squaring, summing and dividing by the theoretical infinite population, we obtain

$$\sigma^2 = \frac{R\sigma^2\sigma^2_r}{r\Sigma^2} + \frac{4R\sigma^2_\sigma(\sqrt{r}-1)^2}{\Sigma^2} + \frac{4R\sigma^2\sigma^2_\Sigma(1-\sqrt{r})^2}{\Sigma^4}$$
$$+ \frac{4R\sigma\sigma_\sigma\sigma_r r_{r\sigma}(\sqrt{r}-1)}{\Sigma^2\sqrt{r}} + \text{(three zero terms . . .)}$$

The "three zero terms . . ." occur because there is no correlation between variables resulting from data from different populations.

By substituting the values of σ_σ, σ_Σ, σ_r, and $r_{r\sigma}$, and simplifying, we obtain

$$\sigma_R^2 = R(1-\sqrt{R})^2\left[\frac{2(N+n)}{nN} + \frac{(1-r)(1+r^2)+2\sqrt{r}(1+\sqrt{r})^2}{nr}\right]$$

whence

$$\sigma_R = (1-\sqrt{R})\sqrt{R}\left[\frac{2(N+n)}{nN} + \frac{(1-r)(1+r^2)+2\sqrt{r}(1+\sqrt{r})^2}{nr}\right]^{1/2} \quad (8)$$

By similar derivation, we obtain for the standard error of expression (6)

$$\sigma_r = (1-\sqrt{r})\sqrt{r}\left[\frac{2(N+n)}{nN} + \frac{(1-R)(1+R^2)+2\sqrt{R}(1+\sqrt{R})2}{NR}\right]^{1/2} \quad (10)$$

Formulas (8) and (10) give the standard errors of expressions (4) and (6) respectively. To avoid imaginary numbers when substituting, we use the absolute values (neglect signs) of the given and estimated coefficients for formulas (8) and (10).

We will now compare some characteristic values predicted by Kelley's formula (1) with corresponding values predicted by formula (2). Some hypothetical data were computed and are summarized in Tables I and II.

The data in Table I and Table II are plotted in Fig. I. Lines (1) are drawn from the data in Table I which resulted from estimating by Kelley's formula (1); and lines (2) are drawn from data in Table II which resulted from estimating by formula (2).

From a study of Fig. I it is seen that the difference between the given coefficient and the coefficient estimated by Kelley's formula (1) is greater than the difference between the given and the estimated coefficient when the estimating is done by formula (2). There are two exceptions—when $\Sigma/\sigma = 1$, and when either coefficient is unity—the last of which is only theoretical. This difference increases for decreasing values of the coefficients. The coefficient estimated by formula (2) is affected less than the coefficient estimated by formula (1) by a unit change in the given coefficient. The range of "Σ/σ-Values" is sufficient to indicate the trends of either lower or higher values of the coefficients and the ratio of the sigmas. The lines for either formula taken separately do not overlap; whereas, there is considerable overlapping when the lines (1) and (2) are compared. This is increasingly true for decreasing values of coefficients, and would indicate that the values estimated by formulas (1) and (2) are not interchangeable. Obviously the standard error values are not interchangeable.

We will now compare formulas (1) and (2) to determine which is the better formula to use when estimating the reliability coefficient. We will define "better" in

TABLE I

Predicted R-Values for Given Values of r and Σ/σ, by Kelley's Formula (1)

Values of r	Σ/σ-Values												
	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0
0.00	—0.56	—0.23	0.00	0.17	0.31	0.41	0.49	0.56	0.61	0.65	0.69	0.72	0.75
0.25	—0.17	0.07	0.25	0.38	0.48	0.56	0.61	0.67	0.71	0.74	0.77	0.79	0.81
0.50	0.22	0.38	0.50	0.59	0.65	0.70	0.74	0.78	0.80	0.83	0.85	0.86	0.88
0.75	0.61	0.69	0.75	0.79	0.83	0.85	0.87	0.89	0.90	0.91	0.92	0.93	0.94
1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Note: This table is read as follows: When $r = 0.50$ and $\Sigma/\sigma = 1.3$, the predicted value of R is 0.70.

TABLE II

Predicted R-Values for Given Values of r and Σ/σ, by Formula (2)

Values of r	Σ/σ-Values												
	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0
0.00	—0.06	—0.01	0.00	0.01	0.03	0.05	0.08	0.11	0.14	0.17	0.20	0.22	0.25
0.25	0.14	0.19	0.25	0.29	0.34	0.38	0.41	0.44	0.47	0.49	0.52	0.54	0.56
0.50	0.41	0.46	0.50	0.54	0.57	0.60	0.63	0.65	0.67	0.69	0.70	0.71	0.72
0.75	0.70	0.73	0.75	0.78	0.80	0.81	0.82	0.83	0.84	0.85	0.86	0.87	0.87
1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Note: This table is read as follows: When $r = 0.50$ and $\Sigma/\sigma = 1.3$, the predicted value of R is 0.60.

terms of the size of the standard error, that is, that formula is the better (other things being equal) which estimates a reliability coefficient with the smaller standard error. Other things are as nearly equal here as could be desired.

The standard error formulas (8) and (10) when compared respectively with the standard error formulas (7) and (9), reveal the fact that the standard errors by (8) and (10) are less than the standard errors by (7) and (9). It is possible for values from (7) and (9) to be more than ten times the values from (8) and (10) when the given data are identical. Formulas (8) and (10) compared with formulas (7) and (9) respectively, show a similarity in that each is the product of two main factors. The values of the first factors are considerably less in formulas (8) and (10) than those in formulas (7) and (9). There is one exception—when the value of the estimated coefficient is unity—in which case all standard errors are zero. The second factors in formulas (8) and (10) and formulas (7) and (9) are functions of $(N+n)/nN$. As the difference between N and n diminishes, that is, as the population sizes become more alike, the value of $(N+n)/nN$ diminishes. The weighing of the numerator $(N+n)$ is greater in formulas (7) and (9) than in formulas (8) and (10), except when the given coefficient has a value of unity. Then they have equal weights. Again all standard error values would be zero. For given coefficient values the second terms in the brackets of formulas (8) and (10) are slightly larger than the difference between the first bracket-terms of (8) and (10) and the bracket-terms of (7) and (9) respectively; but these differences are more than counterbalanced when the standard error values are computed. This is due to the smaller values of the first factors in (8) and (10) which enter into the product. In all practical circumstances the standard error values from formulas (8) and (10) are less (very little in a few cases) than the

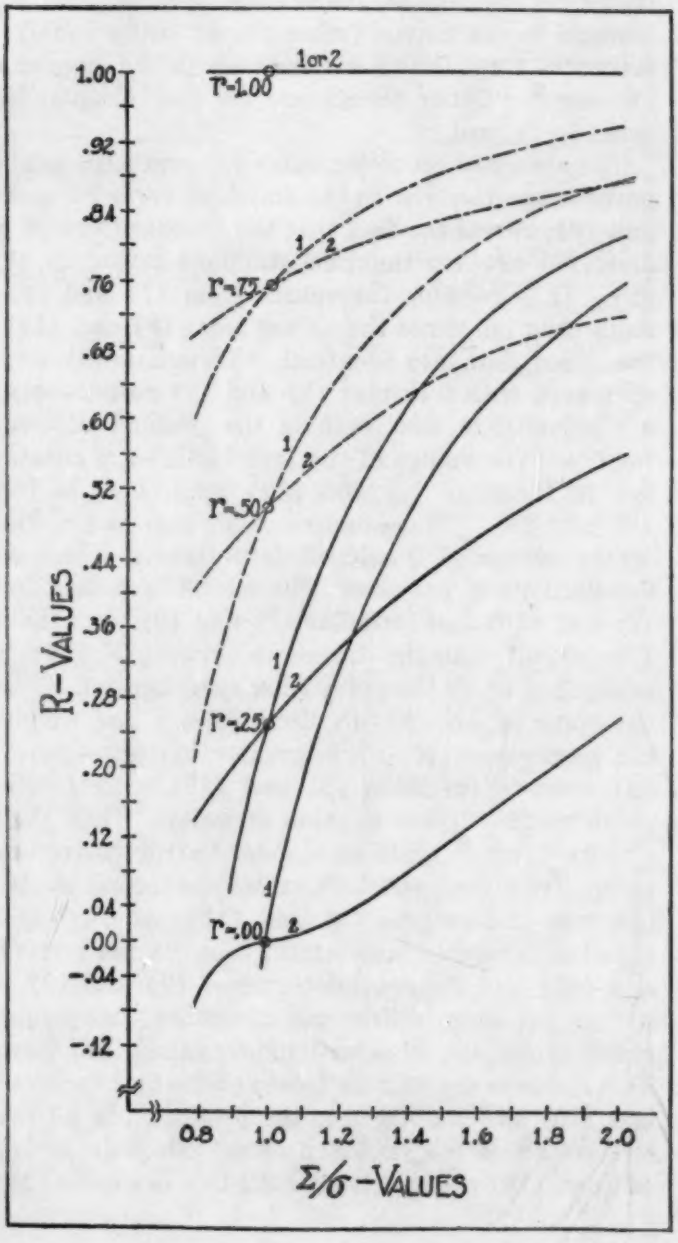


FIGURE I

Showing R-Values Estimated by Formulas (1) and (2) for Certain Given Values of r and Σ/σ

standard error values from (7) and (9) respectively. (Two impractical cases occur: (1) when the given and estimated coefficients are both zero, formulas (8) and (10) reduce to indeterminate forms, (2) when the estimated coefficient is different from zero and the given coefficient is zero, the standard error values of formulas (8) and (10) are meaningless).

By actually substituting in the standard error formulas we can point to this superiority in estimating by formula (2) rather than by formula (1). The data are given in Table III.

From Table III, for given values of R, Σ/σ, and populations (N; n), we see that the estimate made by formula (2) is greatly superior to the estimate made by formula (1) when the coefficient values are small. As the coefficient values more nearly approach unity, this superiority tends to diminish. We may say, however, that the estimate made by formula (2) is superior to that made by formula (1) throughout the entire range. The smaller standard error value is the proof of this statement. (Similar statements are true when r is given and R is estimated.)

In conclusion, when it is necessary to estimate the reliability coefficient of a test from one range of talent, when the reliability is known for another range, the estimate made by formula (2) is superior to that made by formula (1). The comparison of standard error values substantiates this fact.

TABLE III

Showing the Standard Error Values of Coefficients Estimated by both Kelley's Formula (1) and by Formula (2), for Given Values of R, Σ/σ, and Populations (N; n)

Given Data			Estimated Coefficients With Respective Standard Errors		Standard Errors Compared
Reliability Coefficient R	Ratio of Sigmas Σ/σ	Populations N; n	Kelley's $r\pm\sigma$	The Writer's $r\pm\sigma$	Ratio of Standard Errors σ/σ
0.25	1.0	100;100	0.25±.17	0.25±.11	1.5
0.25	1.0	100;50	0.25±.20	0.25±.11	1.8
0.25	1.0	200;50	0.25±.17	0.25±.12	1.4
0.25	1.5	100;100	—0.68±.70	0.06±.08	8.7
0.25	1.5	100;50	—0.68±.83	0.06±.08	10.3
0.25	1.5	200;50	—0.68±.70	0.06±.05	14.0
0.25	2.0	100;100		0.00±.00	---
0.25	2.0	100;50	Meaningless	0.00±.00	---
0.25	2.0	200;50		0.00±.00	---

TABLE III (Continued)

Given Data			Estimated Coefficients With Respective Standard Errors		Standard Errors Compared
Reliability Coefficient R	Ratio of Sigmas Σ/σ	Populations N; n	Kelley's $r \pm \sigma$	The Writer's $r \pm \sigma$	Ratio of Standard Errors σ/σ
0.64	1.0	100;100	0.64±.12	0.64±.06	2.0
0.64	1.0	100;50	0.64±.15	0.64±.06	2.5
0.64	1.0	200;50	0.64±.13	0.64±.07	1.8
0.64	1.5	100;100	0.19±.17	0.49±.08	2.1
0.64	1.5	100;50	0.19±.21	0.49±.09	2.3
0.64	1.5	200;50	0.19±.15	0.49±.07	2.1
0.64	2.0	100;100	—0.44±.32	0.36±.09	3.5
0.64	2.0	100;50	—0.44±.39	0.36±.09	4.3
0.64	2.0	200;50	—0.44±.32	0.36±.08	4.0
0.81	1.0	100;100	0.81±.04	0.81±.02	2.0
0.81	1.0	100;50	0.81±.05	0.81±.02	2.5
0.81	1.0	200;50	0.81±.04	0.81±.02	2.0
0.81	1.5	100;100	0.57±.09	0.72±.03	3.0
0.81	1.5	100;50	0.57±.09	0.72±.03	3.0
0.81	1.5	200;50	0.57±.09	0.72±.02	4.5
0.81	2.0	100;100	0.24±.15	0.64±.03	5.0
0.81	2.0	100;50	0.24±.18	0.64±.04	4.5
0.81	2.0	200;50	0.24±.17	0.64±.04	4.2

TABLE III (Continued)

Given Data			Estimated Coefficients With Respective Standard Errors		Standard Errors Compared
Reliability Coefficient R	Ratio of Sigmas Σ/σ	Populations N; n	Kelley's $r\pm\sigma$	The Writer's $r\pm\sigma$	Ratio of Standard Errors σ/σ
0.85	1.0	100;100	0.85±.03	0.85±.02	1.5
0.85	1.0	100;50	0.85±.04	0.85±.03	1.3
0.85	1.0	200;50	0.85±.03	0.85±.02	1.5
0.85	1.5	100;100	0.66±.07	0.77±.04	1.7
0.85	1.5	100;50	0.66±.10	0.77±.04	2.5
0.85	1.5	200;50	0.66±.08	0.77±.04	2.0
0.85	2.0	100;100	0.40±.12	0.71±.05	2.4
0.85	2.0	100;50	0.40±.15	0.71±.06	2.5
0.85	2.0	200;50	0.40±.12	0.71±.04	3.0
0.90	1.0	100;100	0.90±.02	0.90±.02	1.0*
0.90	1.0	100;50	0.90±.03	0.90±.02	1.5
0.90	1.0	200;50	0.90±.02	0.90±.02	1.0*
0.90	1.5	100;100	0.77±.05	0.86±.03	1.6
0.90	1.5	100;50	0.77±.05	0.86±.03	1.6
0.90	1.5	200;50	0.77±.05	0.86±.02	2.5
0.90	2.0	100;100	0.60±.08	0.81±.04	2.0
0.90	2.0	100;50	0.60±.08	0.81±.04	2.0
0.90	2.0	200;50	0.60±.08	0.81±.02	4.0

TABLE III (Continued)

Given Data			Estimated Coefficients With Respective Standard Errors		Standard Errors Compared
Reliability Coefficient R	Ratio of Sigmas Σ/σ	Populations N; n	Kelley's $r \pm \sigma$	The Writer's $r \pm \sigma$	Ratio of Standard Errors σ/σ
0.95	1.0	100;100	0.95±.01	0.95±.00	---
0.95	1.0	100;50	0.95±.01	0.95±.01	1.0*
0.95	1.0	200;50	0.95±.01	0.95±.01	1.0*
0.95	1.5	100;100	0.89±.02	0.90±.01	2.0
0.95	1.5	100;50	0.89±.02	0.90±.02	1.0*
0.95	1.5	200;50	0.89±.02	0.90±.01	2.0
0.95	2.0	100;100	0.80±.04	0.88±.01	4.0
0.95	2.0	100;50	0.80±.04	0.88±.02	2.0
0.95	2.0	200;50	0.80±.04	0.88±.02	2.0

Note: This table is read as follows: When R is given as 0.25, the ratio of the two sigmas Σ/σ given as 1.0, and the popuation of N and n given as 100 each, the r predicted by Kelley's formula (1) is 0.25±.17 (the standard error computed by formula (9)), and the r predicted by the writer's formula (2) is 0.25±.11 (the standard error computed by formula (10)). The ratio of these two standard errors is 1.5.

* These ratios are due to the rounding off of the decimals to two places. When the decimals are carried to three or more places, the ratios are greater than one.

THE EFFECT OF MOTHERS' OUT-OF-HOME EMPLOYMENT UPON CHILDREN'S IDEAS AND ATTITUDES

SELMA M. MATHEWS
Ohio Wesleyan University

Ever increasing numbers of mothers are seeking employment outside the home. Many of these women desire work for its financial returns while others are impelled to seek out-of-home employment for the social and cultural contacts it supplies and the opportunities it offers to realize more completely certain interests and capacities which they possess. At the same time, the movements for adequate child training, parental education, and sound mental health have grown rapidly.

A question arises in the minds of many women as well as in the minds of those interested in parental education and child welfare as to the effects of out-of-home employment upon the conditions in the home and the child's attitudes toward these conditions. Sound mental health is likely to develop more successfully in the fertile soil of a happy well ordered home life, free from an excess of irregularities, disturbances and flurry. How does the daily absence of mothers from their homes affect such conditions? More directly, what are the differences, particularly the psychological and social differences, between the homes of working mothers and those of non-working mothers? And in the second place what are the differences in the attitudes of children of working and non-working mothers toward these conditions? It was for the purpose of discovering answers to these two questions that the present study was initiated.

I. Collection of Data

The study was begun and carried through from the child's standpoint and, thus, represents only children's attitudes in the light of their own experience and sense of values. By conducting six preliminary studies upon 781 elementary, high school and university pupils, there were secured 5,638 unsigned responses or statements of attitude relative to mother's place in the home and home life situations in general. These children represented a wide variety of types of homes and communities. A questionnaire called a "Home Blank" was constructed from these data after a detailed process of refinement of the children's statements. Subjects were requested to answer the questions as to whether the particular situation existed in their homes and, adjacent to this, whether they were glad or sorry for the answer they had given, which meant glad or sorry for the situation as it existed in their homes. The form of the blank is shown below:

DIRECTIONS

Here are some questions about the most common things which happen in our homes. This is an example:

1. Do you sleep at home?_____ ____ _____

If you do sleep at home, write "yes" on the first line.
If you do not sleep at home, write "no" on this line.
Then, if you are glad for the answer you have given, write "glad" on the *second* line.
If you are sorry for the answer you have given, write "sorry" on the second line.
Now remember: "Yes" or "no" is to be written on the *first* line.
"Glad" or "sorry" is to be written on the *second* line.

Here are a few questions to practice on before going ahead:

2. Do you have a dog?_____ ____ _____
3. Does your mother punish you sometimes?_____ ____ _____
4. Do you have the measles?_____ ____ _____

Now go ahead. Be sure to mark *every* question.

The Home Blank was administered to 568 children of grades five to nine found in school rooms, city playgrounds, boys' camps, and daily vacation Bible schools of nine different communities of various types and sizes in Ohio, Colorado and North Carolina. A free, informal atmosphere was established during the administration of the blanks. Interest devoid of suspicion was secured. Step by step the procedure was explained until all of the children were certain of the reactions expected. They were not unduly rushed but were encouraged to avoid wasting time in thinking over the questions. Meanings of words or phrases were always made clear upon request. In every group a very happy, informal, and cooperative attitude was secured. No signatures were affixed to the papers. The work seemed to be enjoyed by all. The children seemed quite sincere in their efforts to answer questions truthfully to the best of their ability. Many wanted to give additional information on several of the questions. No omissions were permitted unless the pupil had no father or was an only child. In such instances questions pertinent to these factors were omitted. The situation was so arranged that there was no suspicion at any time of the real selective factor in the blank. One boy at a camp was overheard explaining to another that this was "simply a survey to see how people live in the average home."

One hundred of the 568 children to whom the Home Blank was administered were found to represent homes in which the mother worked at a full time job away from the home. These were carefully paired with one hundred papers of children of non-working mothers, keeping constant the age, grade, sex, school, community and father's occupation, using the Sim's Score Card occupational chart. In case of the death of or separation from the father, occupational pairing was made either with similar cases or by consideration of the combined picture of the mother's work and the size of the family and house,

for it generally followed that in case the husband was employed at some menial or unskilled labor, the wife was likewise employed. The remainder of this report will deal only with the replies of those two hundred children and their differences.

The reliability of the Home Blank was checked by several methods and found quite satisfactory for this purpose. The reliability coefficient (self reliability using the Brown formula) for facts was .77 ± .03; for attitudes, .69 ± .03. A key constructed from judgments of adults, as is explained later, was used as a criterion in scoring individual papers for computing reliabilities.

The validity of the items rests upon the extreme care with which they were chosen and upon a comparison of results of the preliminary work with the final study. The work was taken seriously by the children, advantage was taken of the freedom to ask questions, the blanks were unsigned, and the results showed only few possibly bizarre answers. The detail techniques used in the collection of this data may be read elsewhere.[1]

II. Interpretations

The responses of the children were studied in four different ways. First, their answers of "yes" and "no" were interpreted to indicate the existence or non-existence of the specific conditions referred to in the item. The proportion of "yes" responses to all responses for each item is termed a *condition score.*

In the second place, each item was scored in such a way as to determine the extent to which the condition it reflected was approved by the children. A response of "yes—glad" signifies approval of the condition in question. Likewise "No—sorry" indicates the same attitude

[1] Mathews, Selma M.: *The Effect of Mothers' Out-of-Home Employment Upon Children's Ideas and Attitudes.* An unpublished doctor's dissertation, Ohio State University, Columbus, Ohio, 1931.

though negatively expressed. Therefore, all "Yes"—"Glad" and "No"—"Sorry" responses were combined. The proportion this sum is of the total number of responses to the item is termed an *attitude score.*

Thirdly, each child's blank was scored in such a way as to show the extent to which the home conditions indicated on his blank were in agreement with adult social standards and the extent to which his attitudes were in agreement with socially desirable attitudes. Eleven graduate students served as adult judges in establishing standards for home conditions and for socially approved attitudes. These two scores will be termed *socially approved condition score* and *socially approved attitude score.*

Finally, a score was computed on the basis of the number of "Glad" responses only. This score like the others was converted into a percentage by finding the proportion which his "Glad" responses were of the sum of "Glad" and "Sorry" responses. This is termed a *happiness* score.

It will be seen therefore that there is a *condition score* indicating the existing home conditions, an *attitude score* indicating the child's attitudes toward these conditions, a *socially approved condition and attitudes score* indicating the extent to which the conditions in the home and the child's attitudes toward them agree with adult social standards and, finally, a *happiness* score indicating the child's happiness because of or in spite of his conditions. The results in terms of these various scores will be taken up in turn.

A. Condition Scores.

The responses of children of workers and non-workers differ on a large number of items both as to fact and as to attitude. In order to pick out those differences which are significant statistically, the differences and their probable errors were figured for each item. For ex-

ample, notice the first question, the question numbered 2, in Table I, "Does your mother know most of your friends?" Ninety-one per cent of worker children gave the response "Yes" to the question. Seventy-eight per cent of the non-worker children answered "Yes." The ratio of the difference to the P. E. of the difference is 3.8. Only items which produced a ratio of 3 or above were selected as significant enough for comparisons. If the ratio is 3 the results likely would be the same in ninety-eight cases out of every one hundred if the study were repeated. In all of the interpretations which follow, this criterion of statistical significance is used.

Table I includes only those items out of the total one hundred items in the Home Blank which show a statistically significant difference between workers and non-workers in respect to home conditions indicated by the condition scores. (Attitude scores and their differences are also shown in the right hand sections of this table and these will be referred to later.) There are twenty-nine of these items. If the condition score (percentage of "Yeses") for the worker is the larger, the difference is listed in the column headed "W." If that for the non-worker is the larger, the difference is listed in the "N.W" column.

As it will be seen later, the differences in favor of the workers are largely on items not socially approved, while the differences in favor of the non-workers are generally on approved items. In fact, one might write a vivid description of the average home where the mother works, by incorporating into it the twenty-nine items shown in Table I. Unkept clothes, cold meals, abbreviated social life, hurry and bustle, a scolding, tired mother, and so forth, would be included in the picture. But, on the other hand, this mother knows her children's playmates, it is quieter about the home, the children have more freedom to do as they please, and adjustments of various types are made so that the mother does not do double

TABLE I

Statistically Significant Conditions

Item No.	Statement of Item	Conditions					Attitudes				
		% of "Yes" Responses		Diff. in Favor of		Ratio Diff. to P.E. Diff.	% of Approval		Diff. in Favor of		Ratio Diff. to P.E. Diff.
		W.	N.W.	W.	N.W.		W.	N.W.	W.	N.W.	
2.	Does your mother know most of your friends?	91	78	13		3.8	96	94	2		1.0
24.	Do you often have to wear soiled clothes to school?	30	3	27		8.2	2	2			
32.	Does Mother often interrupt you in your play?	55	41	14		3.0	25	11	14		3.6
48.	Do you often cook at home yourself?	63	42	21		4.6	73	63	10		2.3
54.	Do you often do as you please when your mother is gone?	41	27	14		3.3	37	28	9		2.0
59.	Do your folks usually send your laundry out?	38	23	15		3.5	59	42	17		3.6
60.	Does Mother sometimes leave you children at home alone?	61	46	15		3.1	39	28	11		2.4
68.	Does Mother go off to work in the day time?	98	3	95		63.3	40	5	35		9.7
73.	Do you usually have to hurry around home?	52	33	19		4.1	38	29	9		2.0
83.	Does your Mother often scold you children when she is tired?	74	45	29		6.4	20	14	6		1.7
91.	Do you sometimes have to make your own breakfast?	45	24	41		4.3	40	23	17		4.0
3.	Does your Mother usually see you off to school?	49	80		31	7.2	88	90		2	.7
5.	Does your Father sometimes play games with you?	65	80		15	3.4	90	94		4	1.5
9.	Are your stockings usually kept darned?	88	97		9	3.6	95	96		1	.5
13.	Does your Mother sometimes visit the neighbors?	75	91		16	4.6	87	96		9	3.3
23.	Do you usually have a hot cooked meal at noon?	75	93		18	5.3	96	96			
33.	Does your father have a job?	84	94		10	3.0	99	100		1	1.1
34.	Does Mother usually take care of the sick children in your home?	93	99		6	3.2	91	97		6	2.6
35.	Is Mother at home most of the time?	27	95		68	20.0	89	93		4	2.7
36.	Are your meals usually on time?	92	99		7	3.7	98	100		2	2.2
40.	Is it often noisy at home?	25	35		10	3.0	8	9		1	.4
44.	Does Father sometimes help take care of the children?	68	83		15	3.2	83	80	3		1.0
50.	Is somebody usually home with you after school?	73	89		16	4.4	86	90		4	1.3
58.	Does Mother do all the cooking at home?	23	58		35	8.1	40	57		17	3.6
71.	Do you have nice furniture?	88	96		8	3.2	96	99		3	2.0
76.	Does Mother do most of the house-work?	41	85		44	10.7	32	55		23	5.0
80.	Is Mother usually at home after school?	42	97		45	15.3	86	99		13	5.2
89.	Is your home a cozy one?	88	96		8	3.1	96	97		1	.6
94.	Do you usually have hot cooked suppers?	95	100		5	3.3	100	100			

duty in the way of all housework, cooking, and laundry in addition to her outside occupation.

It is very essential to consider the size of the percentages before interpreting any of the several items. Notice item 24 for example. Although the difference between the worker and non-worker percentage scores is large and significant, these percentages are comparatively small, so that these figures must be interpreted as characterizing a minority of the workers' children. In other words, 70 per cent of workers' children do not suffer the indignity of having to wear soiled clothes to school. Question 36 shows high percentage for both workers and non-workers. Again, although there is a statistically significant difference here, it must be kept in mind that 92 per cent of the workers have their meals on time. The 8 per cent who suffer irregular meals is a small number out of one hundred families.

B. Attitude Scores.

Attention will now be directed to the attitude scores and their differences for the two groups, children of working and non-working mothers.

In interpreting these scores it must be remembered that they represent the extent to which the children approve the home conditions indicated in the items. The "Yes"—"Glad" and "No"—"Sorry" responses were interpreted as showing the same attitude and were combined in computing the percentages shown as attitude scores.

A casual inspection of the attitude scores for these two groups in the right hand portion of Table I will show the great similarity which exists between the attitudes of the two groups. Conditions which are highly approved by one group are generally approved by the other. It is evident from inspection that there is greater similarity between the two groups in respect to attitudes than in respect to home conditions.

With only one exception, the workers' children approve more strongly than non-workers' children those conditions which exist more frequently in their homes. The attitudes of both groups toward wearing soiled clothes (Item 24) are the same. With only three exceptions the workers' children approve less strongly those conditions which occur less frequently in their homes. These exceptions are items 23, 94 and 44. Although they have hot cooked meals at noon and hot cooked suppers less frequently than non-workers' children, they favor these conditions to the same extent. And, although their fathers help take care of the children less frequently than fathers of the control group, they would be more pleased to have him do so. This latter difference may be due to the fact that the children of working mothers probably have to share in taking care of younger brothers and sisters more frequently than children of non-working mothers and this would be one means of relief.

Since the differences in favor of the worker group are, on the average, greater than those in favor of the non-worker group this would tend to show that the workers' children are more critical of actual home conditions than the children of non-workers. In other words, if each group could change its home conditions to correspond with its indicated attitudes there would be greater changes in the homes of working mothers.

Of the twenty-nine items in Table I only nine show significant differences in respect to attitudes. These nine have been brought together in Table II. The condition scores for these items are listed for comparison. The attitude scores for the worker group are the larger on the first five items of this table and those of the non-worker group are larger on the last four items.

The study itself furnishes no evidence which shows the causes of these differences. It may be noticed, however, that the activities indicated by the four items less highly approved by children of working mothers are such as

might interfere with the mother's work or be an added burden to her. To visit one's neighbors, be at home after school, do all the cooking, and do most of the housework would cut in to the working time of many employed mothers. It may be that a certain number of children marked these items "No"—"Glad" either out of thoughtfulness for the mother or out of a desire to have more time and attention than a part-time mother could give. A comparison of condition scores shows that these activities are actually performed less frequently by the working mothers.

TABLE II

Statistically Significant Attitudes

Item No.	Statement of Item	Condition Scores % of 'Yes' Responses		Attitude Scores % of Approval		Diff. in Favor of		Ratio Diff. P. E. Diff.
		W.	N.W.	W.	N.W.	W.	N.W.	
92.	Are your clothes nice?	96	95	100	96	4		3.1
59.	Do your folks usually send your laundry out?	38	23	59	42	17		3.6
91.	Do you sometimes have to make your own breakfast?	45	24	40	23	17		4.0
68.	Does Mother go off to work in the daytime?	98	3	40	5	35		9.7
32.	Does Mother often interrupt you in your play?	55	41	25	11	14		3.6
13.	Does your Mother sometimes visit the neighbors?	75	91	87	96		9	3.3
80.	Is Mother usually at home after school?	42	97	86	99		13	5.2
58.	Does Mother do all the cooking at home?	23	58	40	57		17	3.6
76.	Does Mother do most of the housework?	41	85	32	55		23	50

Comparison of Normal and Broken Homes.—As there were thirty-one cases among the worker group in which the father was deceased and seven cases in which the father and mother were separated, it was thought that perhaps this factor of an abnormal home situation might affect the children's attitudes toward the mother's out-of-the-home occupation. The thirty-eight cases were grouped together, thus making a group of eighteen girls

TABLE III

Responses to Items 35 and 68 with Differences Between Groups in Percentages

Item No.	Condition or "Yes" Responses				Attitude Responses			
	Worker	Non-Worker	Diff. in Favor of W.	Diff. in Favor of N.W.	Worker	Non-Worker	Diff. in Favor of W.	Diff. in Favor of N.W.
35	27	95		68	89	93		4
68	98	3	95		40	5	35	

and twenty boys represented. These thirty-eight were paired with eighteen other girls and twenty other boys of working mothers whose husbands were living with the family and working. The paired cases were chosen at random, although pairs were kept within the same school and grade.

The condition score for the normal family group was 75.5 and for the broken home group 69.7. The attitude scores were 85 and 84, respectively. The difference between the first pair of scores is significant while the latter one is not, although the broken home group ranked lower in respect both to home conditions and attitudes.

Children's Attitudes Toward Mother's Out-of-Home Employment.—Do children approve out-of-home employment for their mothers? Let us examine their replies to the two questions which were inserted in the Home

Blank in order to get this attitude. Question 35 asked whether the mother was at home most of the time, while question 68 asked whether she went off to work in the daytime. Table III answers this question quite clearly.

The writer learned that a good many of the mothers returned from their work just before or at the same time that school was out. Therefore, to 27 per cent of these children it would seem that mother was at home most of the time as her absence from the home was largely during the time they also were absent. There were very high percentages of approval on the item concerning mothers being at home most of the time. For the worker group it figured 89 per cent and for the non-worker group, 93 per cent. Although the workers' children express their approval less forcefully, nevertheless, they rank very close to the other group on this attitude. The returns on the second item were still more direct. Only 40 per cent of the workers' children favor having their mothers go to work in the day time. The 40 per cent presents a loyal difference in comparison with the 5 per cent of the non-workers' children, but it must be concluded from these figures that they are in the minority. Speaking generally one could not strongly defend mothers in their outside employment if this one item were to decide the question. However, this would not be fair either to the mothers or the children. Significant modifications of this attitude are shown when other items are taken into consideration and a general attitude is considered.

In a large number of cases, the children suggested that it would be of great advantage to have mother off at work so there would be less parental suppression, domination, and interruptions around home. One of the strongest criticisms against the mother was her continuous interruptions and housework noises so that the child could not think, play, study, or even "live in peace." This type of factor was referred to more than any other

in the preliminary papers of other children from which the items of the Home Blank were constructed. They expressed the desire to be rid of mothers so they could do more as they pleased, all the way from "feeding the cat more" and "blowing horns" and "playing with noisy motors" in the house to "having the gang in often," "having the dog in the house," and "making candy every day."

C. Socially Approved Conditions and Attitudes.

To what extent were these children's home conditions and attitudes in agreement with adult judgments of ideal conditions and attitudes? Eleven graduate students acted as judges in establishing a key against which to check the papers for such agreement. They were instructed to check all questions "yes" which represented the ideal or socially approved home situation. All other questions were to be answered by "no." Answering on this basis automatically made "glad" the only correct attitude response on each of the items. The decision was practically unanimous in disfavor of twenty-six of the items. On four additional items there was more controversy, but these were included as disapproved because they were in disfavor by over half of the judges. There were only a few scattering disapprovals of some of the remaining items. Of these thirty disapproved items, fifteen occurred in the first half of the blank and fifteen in the last half.

The papers of the children of both groups were scored according to this key. Table IV shows in terms of percentages the extent to which each group agrees with the adult standards and how they compare with each other.

It will be seen in Table IV that non-workers' children live under home conditions more favorable from the standpoint of adult standards than workers' children. Their attitudes also are slightly more in agreement with these standards. The variability of scores for the worker

group is somewhat larger than for the non-worker group as is shown by the larger range and quartile deviations. This means that the extent of agreement with adult standards varied more among children of working mothers.

It is conceivable by this method of scoring that the workers' children were somewhat penalized for attitudes which were psychologically good, but from a social standpoint undesirable. To illustrate: Suppose a child

TABLE IV

Socially Approved Scores in Terms of Percentages for Each Group

Group	Socially Approved Scores	
	For Conditions	For Attitudes
Workers' Children	72 range = 52-91 Q = 11.5	83.8 range = 48-96 Q = 9.1
Non-Workers' Children	77.5 range = 55-92 Q = 10.4	85.6 range = 66-96 Q = 6.9

had stated that he did not usually have a hot cooked meal at noon, and he was glad. The judges would have penalized the condition because a warm meal is more desirable and they would have deducted for the attitude because one should be, theoretically, sorry for the less desirable. If conditions in the home were such, however, that a hot meal at noon was out of the question, it is psychologically a good thing for the child to be satisfied or glad. In other words adults insist that ideally, children should live in certain conditions and be glad for them, but if children with less favorable conditions are able to be happy in spite of them this is bound to make for happier family relations and, therefore, to result in fewer maladjustments in the face of the realities of life.

Many of the interviews gave the writer this impression. A typical one is cited. One poor, ragged, little fellow was interviewed on a city playground while his mother, who could neither read nor write, was at work in a hosiery mill. His general bearing gave the impression that he had not a care or moment of unhappiness. When generally attractive questions were asked, he smiled and, without a single sign of scorn or wistfulness, answered truthfully, although his answers were not desirable ones. Such things as nice clothes, parties, nice furniture, pretty things, spending money, and trips made only a distant appeal to him. His responses often included some reference to the fact that his home was just like all ordinary homes in these respects, or "Whatever is, is right." His father played horse shoes with him and his sister helped him with his lessons. What else could one desire?

In order to check on possible sex differences in attitudes towards the mother's employment, the socially approved attitude score was figured for boys and girls separately. Only the one hundred workers' children were thus scored, there being forty-nine boys and forty-one girls. The median score expressed in percentages for boys was 82.9 and that for girls was 86.8. This gave a ratio of the difference to the P. E. Diff. of 8.0 (using forty-nine hundred and fifty-one hundred as "n"), which is statistically significant in favor of the girls. The girls thus possess a higher degree of approved attitude toward home life situations herein studied. This tends to substantiate the findings of the Character Education Inquiry. Here it was found that girls were more sensitive to both conventional and ideal social standards than boys.

D. Happiness Scores.

A happiness score was computed for each child of the experimental and control groups. This was done on the

basis of all responses which were "Glad" because of good home conditions or in spite of poor ones. The proportion of his "Glad" responses to the sum of his "Glad" and "Sorry" responses determined this percentage.

Table V gives the results of these happiness scores figured for the two groups, in terms of medians.

TABLE V

Happiness Scores in Percentages

Group	Median	Range	Q
Workers' Children	83.7	57-99	11.1
Non-Workers' Children	87.9	65-99	10.1

Note that in the control and experimental groups of one hundred cases each (non-workers and workers, respectively) the median happiness scores are fairly similar. If one were to consider each item of the Home Blank individually he would conclude, probably, that there was no significant difference between the happiness of the two groups. But as Buckingham[2] has so aptly stated, we too often fail to take note of the accumulation of minute advantages which, brought together, are likely to present a more truthful picture of the whole. Accordingly, when the minute differences indicated by these children in respect to happiness were summarized, it was found that the average difference of 4.2 was significant, having a probable error of 1.32. However, the fact that the workers' children rate fairly high in this respect would seem to indicate that from the standpoint of mental hygiene they are well adjusted to their conditions of living. They are fairly contented and happy children.

[2] Buckingham, B. R., "The Accumulation of Minute Advantages," *Journal of Educational Research*, XVI (1927), pp. 136-138.

It might be interesting to define objectively this concept of a happy home in children's terms. Only items from the Home Blank which showed a socially approved condition (as rated by eleven adult judges) to occur in eighty-five per cent or more of the homes was taken as a factor representative of these homes. The assumption was made that if a condition occurred in homes to such a large extent, it surely would need to be included in any description of home conditions which would claim to portray a fairly constant picture of those homes. Thirty-six items were found by means of this selective process. These items describe the conditions herein investigated which go to make up a happy home as children see it and testify to it. The items evaluated on the questionnaire were recast into positive statements in order to interpret more clearly the responses made by the children. It follows that these items or factors are the leading ones, according to this study, which make for the contentment of children in their homes and therefore endear it to them. The revised thirty-six items are listed below:

My home is a comfortable one.
My stockings are usually kept darned.
Our house is usually tidy.
Our home is a happy place most of the time.
Mother usually reads the newspaper.
Our beds are made up every day.
Father and mother usually have good times together.
I do not loiter down town a great deal.
Our meals are usually on time.
Mother usually takes care of the sick children in our home.
Our home is not untidy most of the time.
My parents sometimes have visitors in the home.
We have pretty things in our house.
Home is a place where I am usually proud to invite my friends.
I do not usually have to stay in the house to play.
Home is the pleasantest place in the world.
Mother usually trusts us children.
Mother often settles our disputes.

I am sometimes allowed to buy candy with my own money.
We have nice furniture.
All the members of our family usually eat meals together.
My clothes are usually kept mended.
Mother is usually tired in the evening.
Our home is usually kept in order.
I sometimes have spending money.
It is usually peaceful at home.
We do not usually have cold suppers.
Mother sometimes has time to read.
Our home usually is a cheerful place.
Our noon meals are not usually cold.
Our home is a cozy one.
Mother and I are interested in many of the same things.
My clothes are nice.
We usually have hot cooked suppers.
Mother is usually interested in hearing about our school affairs.
Mother usually helps me out in my troubles.

III. Summary and Conclusions

An attempt has been made to discover children's ideas of and attitudes toward conditions caused by the mother's out-of-home employment. From six preliminary studies, responses were obtained from which a Home Blank questionnaire of one hundred items was constructed. It's validity and reliability were carefully checked. It was so arranged that both home conditions and children's attitudes toward them could be indicated. The blank was administered under standardized conditions to 568 children of grades five to nine, inclusive. One hundred papers of children whose mothers worked full time away from home (the experimental group) were carefully paired with an equal number from children of non-working mothers (the control group).

It is recognized that the numerous factors involved in such a study as this and the relatively small number of cases employed makes one skeptical about stating conclusions with finality. Strictly speaking, proof that a parent's behavior (in this case the mother's out-of-home

employment) is wholly responsible for conditions in the home or the attitudes of the child toward them cannot be furnished. It is probable that other factors in the school and community working in conjunction with the experimental one were not wholly controlled or allowed for. Children's personalities vary so greatly that even siblings react differently to the same treatment from parents. For these reasons, such of the following statements as deal with interpretations will be considered more as inferences than as proven conclusions. The reader will need to decide whether the inferences drawn are justified.

1. Out of the one hundred items in the Home Blank, twenty-nine condition items were found to be statistically significant, while nine attitude items were significant. In each case, the disadvantage fell on the worker side.

2. From the psychological standpoint this study has shown that attitudes are very closely interrelated with environment. With but few exceptions the differences in attitudes between the two groups are in the same direction as the differences in their home conditions. Perhaps this is a corroboration of Professor Symond's statement that, "We like those things toward which we respond positively, and we dislike those things toward which we respond negatively."

3. Although there is a significant difference as to home conditions between normal and broken homes of workers, there is little difference as to the attitudes within the two types of homes.

4. Only forty per cent of the workers' children show approval toward mother's out-of-home employment. At the same time eighty-nine per cent approve of her staying at home most of the time.

5. According to the combined judgments of eleven adult judges, the home conditions as rated by the children found in the workers' homes were not as desirable as those in the non-workers' homes, although this difference was not very great.

6. The attitudes of the worker group were found to be less desirable than those of the non-worker group when measured by adult judgments. However, the similarity of attitudes of the two groups was striking.

7. The girls of working mothers were found to possess more desirable attitudes than the boys. The term desirable refers to the type of attitudes which had been approved by adults.

8. Although the difference is small, it seems probably true to state that children of workers are significantly less happy than those of non-workers.

It should be made clear that it is not contended that this study proves that children do not suffer because mothers work. Another type of study would be necessary in order to answer this problem. The greatest danger of misinterpretation lies in the fact that the Home Blank secures children's evaluations of their home conditions rather than knowledge of actual conditions. For example, one fifth grade child said, "Sure, and I don't have to wear soiled clothes to school. I get a clean shirt every week." This shows the relative meanings attached to these items. Thus, the conclusions which relate to conditions are probably less valid than those which relate to attitudes.

On the whole, the findings of this study are of a very wholesome nature. Statistically speaking, figures show significant differences between the homes of workers and non-workers for both conditions and attitudes, with a slight accumulation of disadvantage to the worker group. But the most impressive thought to be gained from a

study of these figures is that the differences are small. It is beyond the province of this study to prove why and where these attitudes come into being, but it is a fact that the worker's child is not embittered by his condition. He is not a sad, oppressed, unhappy child, suffering greatly, according to his own opinions, because his mother works. This is fortunate. In case the death of the father, economic conditions, or the mother's interests draw her out of the home to work, her children are able to make the necessary adjustments which do not disrupt the family life and the attitudes toward it. If this were not true, this unusual move of the mother likely would change the attitudes of the children toward the mother and home life so that there would result many, many more mental conflicts than now exist.

Possibly an analogous situation may be found in the case of persons who are blind. Those who are suddenly struck with adventitious blindness certainly miss the beauties which they once could see, while those who are congenitally blind do not know what to miss because they are innocent of what others are enjoying. These people are content and happy with their lives for they know no other way.

MENTAL GROWTH OF THE SOUTHWESTERN INDIAN

B. F. HAUGHT
University of New Mexico

INTRODUCTION

Research on the mental equipment of the American Indian has received intermittent attention for the past forty years. Among the topics that have been subjected to experimental investigation are the following: reaction time, sensory acuity, handedness, color preference, community of ideas, musical talent, will-temperament, fatigue and work curves, and intelligence.[1]

This study differs considerably from any of those previously made in that it attempts to show systematically the rate of mental development of the Southwestern Indian from six to sixteen years of age.

INTERROGATIONS INVOLVED

The data presented in this report should throw some light on the following questions: (1) Does the mental growth of the Southwestern Indian parallel the change in his chronological age? (2) Are there periods in the mental growth of the Southwestern Indian that are characterized by acceleration or retardation? (3) How does the Southwestern Indian compare in mental ability with the white race? (4) At what age does the mental

[1] References are omitted since all of them would make a list too long to publish. The writer has accumulated a bibliography of twenty-eight researches on the American Indian, and it is probably not complete.

growth of the Southwestern Indian terminate? (5) To what extent is a test made for the white race indicative of the actual mental equipment of the Southwestern Indian?

PROCEDURE

Tests were administered to Indian sub-adults of the Southwest enrolled in Federal Indian pueblo schools of New Mexico and in the United States Indian schools situated in Albuquerque and Santa Fe.[2] The subjects were all attending school and had been attending for a period ranging from six months for the younger pupils to ten years for the older. It was planned to obtain 100 cases for each chronological age-year from six to sixteen, but on account of the difficulty of obtaining the younger subjects it was necessary to abandon this objective.

[2] About one-half of the tests were administered by Forrest Adams, several by Ralf Gould, both graduate students in the University of New Mexico, and the rest by other graduate students and the writer.

The Pintner-Cunningham Primary Mental Test was used for the younger subjects, the National Intelligence Tests for the intermediate, and the Terman Group Test of Mental Ability for the older. All scores were converted into mental ages.

THE DATA

Table I shows the number of subjects tested, the range of chronological ages, the median chronological age, the median mental age, and the median intelligence quotient for each age-group.

Figure 1 presents graphically a comparison of the mental growth of the Southwestern Indian with the normal mental growth of the white.

TABLE I

Chronological Ages, Mental Ages, and Intelligence Quotients of Southwestern Indians

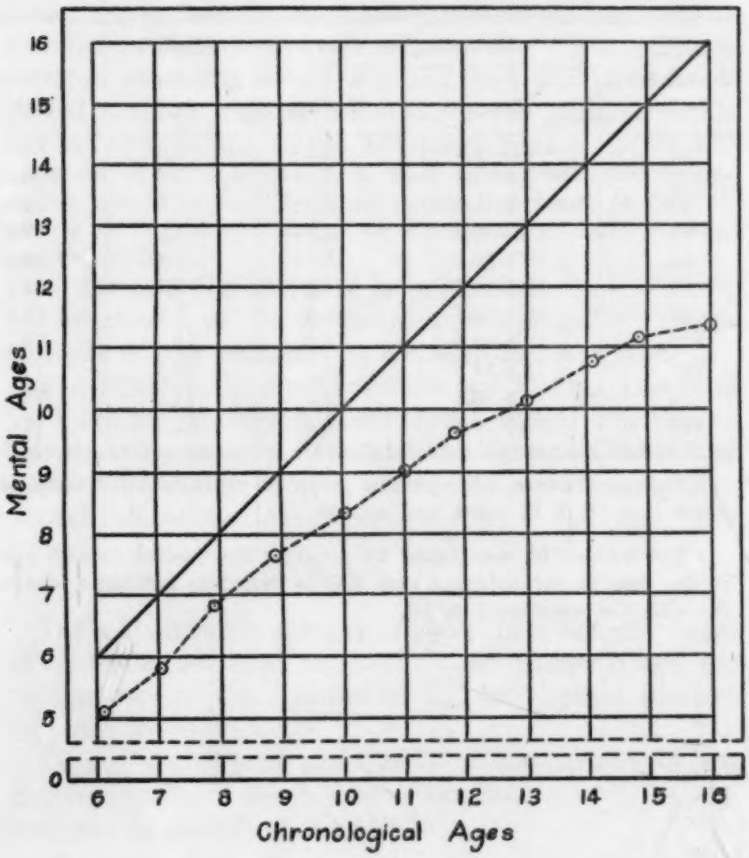
Number Tested	Chronological Ages		Median Mental Age	Median IQ[4]
	Range	Median		
54	5.5- 6.5[3]	6.1	5.1	84
59	6.5- 7.5	7.0	5.8	83
67	7.5- 8.5	7.9	6.7	85
76	8.5- 9.5	8.9	7.7	87
84	9.5-10.5	10.1	8.4	83
92	10.5-11.5	11.0	9.0	82
98	11.5-12.5	11.9	9.6	81
102	12.5-13.5	13.0	10.1	78
107	13.5-14.5	14.1	10.9	77
113	14.5-15.5	14.9	11.2	75
109	15.5-16.5	16.1	11.4	71

[3] For convenience, all ages are given in years and fractions of years instead of in years and months.

[4] The median IQ was found by dividing the median mental age by the median chronological age, and is therefore not quite identical with the usual median IQ.

FIGURE 1

Mental Age Growth of Southwestern Indians Compared With Normal Mental Age Growth of Whites. Whole Line Represents Normal Mental Age Growth of Whites, and Broken Line Mental Age Growth of Southwestern Indians.



CONCLUSIONS

On the basis of the data presented in this study the following conclusions have been derived:

(1) From the chronological ages of six to nine years the mental growth of the Southwestern Indian is approximately parallel with his chronological age but about one year retarded. From the age of nine to sixteen, the retardation of the mental age increases from year to year, and somewhat uniformly.

(2) There are no periods of acceleration in the mental growth of the Southwestern Indian up to the age of sixteen. Retardation is constant from the ages of six to nine; increases from 1.2 years at the age of nine to 2.3 years at the age of twelve; and increases from 2.3 years at the age of twelve to 4.7 years at the age of sixteen.

(3) In mental age, the Indian of the Southwest is from 1.0 to 4.7 years below the whites. His intelligence quotient ranges from 71 to 87. It is highest at nine and lowest at sixteen.

(4) The data do not reveal the age at which mental growth ceases. The increase in mental age, however, is very slight between the ages of fifteen and sixteen.

(5) In terms of scores on the intelligence tests the Indian is below the whites. Naturally the question arises: Is he low in native ability or do the tests fail to measure him accurately because his environment has been different from that of the whites? The evidence that he is low in native ability and not low because he is handicapped by his environment is as follows:

(a) The subjects tested have been in schools from one-half year to ten years. The older the subject the longer he has been in school. It is reasonable to assume that attendance at school should make his environment

and past experience more like that of the whites. Consequently, his ability to take the tests should increase. But the mental growth curve in Figure 1 shows that the older he becomes and the longer he attends school, the farther he falls below the normal mental age for whites. This evidence assumes that no selective factor is involved, and it is likely that selection would operate to increase rather than to decrease intelligence.

(b) The mental growth curve of the Southwestern Indian is about what one would get by observing, over a comparable period, the mental growth of white children, who at the age of six are one year below normal. The average intelligence quotients would remain fairly constant from six to ten or twelve and would then fall off slightly from there on because some of the children begin to reach their mental maturity. An examination of column five in Table I shows that the intelligence quotients of Southwestern Indians remain fairly constant from the ages of six to ten, and from there on decrease.

The findings of this study then present two bits of evidence that the Indians make lower scores than the whites because they are lower in native ability.

THE LANGUAGE LEARNING OF A CHILD

WILLIAM NORWOOD BRIGANCE
Wabash College

The following is a study of the language learning of my daughter, Shirley Jane. It covers three of the important phases of this process—(1) The growth of the vocabulary, (2) a comparison of the parts of speech found in this vocabulary, and (3) growth of the ability to master the phonetic sounds of the English language.

(1) The Growth of the Vocabulary

The first appearance of actual speech—that is, of definite and permanent association of specific speech sounds with specific objects—appeared between the eighth and ninth months. By the ninth month there were four definite words in common use—*daddy, mama, pretty,* and *bye-bye,*—all of which had been acquired within this one month. The next three months, however, instead of bringing an acceleration in this rate of acquisition, as might be expected, produced only three additional words —so that at the end of the 12th month the total vocabulary was seven words. This slowing up, I attribute to what Pelsma[1] has called the "plateau" in learning, which in this instance was produced by SJ making her first attempts at walking. These attempts at walking began before the tenth month, and the first step was taken at eleven months. At twelve months walking was becoming somewhat of an autonomic process.

Following this mastery of the walking process the

[1] Pelsma, J. R., "A Child's Vocabulary," *Ped. Sem.*, 1910, Vol. 17; p. 354.

curve of vocabulary rises sharply—to seventy-five words at fifteen months. But between the fifteenth and eighteenth months there was again a plateau which apparently was a settling down process wherein the mastery of acquired words is perceptibly increased but in which few new words are added. Only twenty-one new words were acquired in this period.

Now comes a period of growth which viewed in the light of previous progress is most remarkable; in the six months between the eighteenth and twenty-fourth months the active vocabulary increased from 96 words to 632! This unusual growth I attribute to two factors: First, SJ was taken upon a long trip during the early part of this period from her home in Indiana through the Rocky Mountains of Colorado and Utah. This widening of experience naturally brought new words for new concepts—*mountains, sheep, turkeys, fish,* etc. Secondly, about this time she developed an avid interest in rhymes, jingles, and Mother Goose stories. These stories became her world of fancy. To the neglect of toys and even food she thumbed the pages of children's books and demanded to be read from them. One can get some idea of this interest when I say that by her second birthday there were by careful test exactly forty rhymes—as Simple Simon, Five Little Pigs, Old Mother Hubbard, Jack and Jill,—of which she could repeat at least three-fourths of the whole rhyme, verbatim and without prompting. This brought a host of new words into articulation, many of which were ushered into definite concept and made a part of the general vocabulary.

In addition to this active vocabulary of 632 words at 2 years, SJ also articulated 140 words borrowed from Mother Goose rhymes. These words also were recognized when spoken by others but careful testing showed that none of them in themselves had been formed into definite concepts. I take the word *black* as an illustration. This word was known from the rhyme, "Baa, baa,

black sheep." When SJ was first shown an actual sheep, she immediately inquired, "That a *black* sheep?" Also the mention of a "black horse," would bring up the concept "black sheep." But the word alone had no concept. The color *black* was not recognized. Since the concept, not the mere articulation, of any word is the ultimate test, this group of words was not included as a part of the vocabulary.

Between SJ's second and third birthday there set in what Stern has called the "complex sentence" stage. Questions of *what* had appeared by the fifteenth month; questions of *where*, by the twenty-fourth month. Questions of *why*, revealing concepts in relationship of ideas and involving complex sentences, were begun at the twenty-seventh month. Questions of *when* appeared shortly after, revealing the development of concepts in time. By the third birthday the total number of words in SJ's vocabulary was 1465.

This vocabulary growth then may be summarized as follows:

Age (in months)	Words
9	4
12	7
15	75
18	96
24	632
36	1,465

Waddle[2] offers the size of a child's vocabulary in relation to others of the same age as a prominent means of judging a child's intelligence. Without question this factor forms an important basis of judgment. Yet the study of SJ's language learning suggests that it is a factor that will bear considerable caution in handling. With her, vocabulary growth received an artificial stimulus from rhyme recitation. Had her interest directed

[2] Waddle, C. W., *Introduction to Child Psychology*, p. 178.

itself toward dolls, toys, or pets before the age of two years she might have acquired the same number of perceptions as she did with rhymes. But a larger proportion of these perceptions would have been unverbalized and the vocabulary correspondingly smaller. Not the vocabulary alone but the total perceptions of the child, verbalized and unverbalized, is the better index of intelligence.

(2) A Comparison of the Parts of Speech

The parts of speech used in a child's vocabulary probably are less significant than was formerly supposed. In the first place the commonest basis of comparison has been between the parts of speech used by children and those used by adults, yet there is little really definitely known about the relative parts of speech used by adults—and what is known suggests that there is less difference between adults and children that was once supposed. Second, and far more important, is the fact that while a child may *speak* a lone word, he *thinks* a whole idea which an adult would express in a sentence. In other words, a child's noun serves for noun, verb, and modifiers. A comparison of parts of speech, therefore, does not give an index of perception. Such a comparison, however, may serve as a basis of study with other child vocabularies and so is here briefly given.

Of the vocabulary at the end of twelve months, four words were nouns, two were interjections, one was an adjective. Subsequent comparisons are shown in the following table:

Age (in months)	Nouns Percentage	Verbs Percentage	Others Percentage
15	76	1	23
18	74	5	21
24	61	20	19
36	58¼	24¼	17½

At once it is noticeable that the proportion of verbs is sharply increasing and that the proportion of nouns

and other parts of speech (chiefly adjectives, adverbs, and prepositions) is declining. This flows inevitably from the fact that at fifteen months SJ was in the "naming" stage, chiefly requiring single words; that at twenty-four months she was well into the "simple sentence" stage, using simple verbs; and that at thirty-six months she was well into the "complex sentence" stage and beginning the use of compound verbs.

(3) Development of the Ability to Master Phonetic Sounds

The acquisition of pure vowel sounds appears to be so easy in the child's learning process as to assume no problem. Diphthongs are more difficult perhaps, although as to this it is hard to speak definitely. The unaided ear at best is a most doubtful instrument for measuring the learning of vowel sounds—and in this case no effort shall be made at measuring the acquisition of vowel sounds except in their appearance as initial sounds.

The mastery of phonetic sounds centers chiefly upon consonants and these shall be treated at some length.

At Twelve Months

At twelve months five initial consonant sounds had appeared—*p, b, m, h* and *d.* In two syllable words the second syllable was a simple repetition of the first—*mama, daddy,* ['da'da], *bye-bye.* The one exception was *pretty* pronounced as ['pɪ'tɪ] in which the vowel was repeated but a new consonant was inserted.

There were no initial vowel sounds.

This acquisition of initial sounds follows the invariable order of normal development, in which the sounds of *p, b,* and *m,* appear first. It has been suggested that the appearance of these sounds first arises from the fact that the child can *see* these sounds being formed

by those around him. This explanation rests upon the assumption that an infant possesses not merely powers of imitation but also that he possesses powers of apprehension to a degree not found in his other behavior. The true explanation would seem to be much simpler and more basic. The sounds of *p, b,* and *m,* are *bi-labial* sounds. In fact, they include all of the pure bi-labial sounds of English. *And the act of suckling has developed in the infant those facial muscles necessary for the utterance of bi-labial sounds.* In short, physiological development—and not apprehension and imitation—forms the most probable basis for the appearance of these sounds before others.

At Fifteen Months

At fifteen months, the following additional initial consonants were well established—*w, k, n, t, g,* and *l,*—making eleven initial consonant sounds out of the total of twenty-two which appear in the English language, or exactly one-half. With one exception these consonants appeared in the traditional order of child learning.[3]

This one exception, however, calls for comment. I refer to the appearance of the initial *l.* This is unusual for two reasons: First, *l,* is supposed according to Tracy to be one of the last sounds to be acquired in normal language learning whereas here it appears among the first half. Second, it was unusually well established at fifteen months—as in *light, lady, Lindy, little.* I venture to question Tracy's contention that the initial *l* is one of the last consonants to be acquired, not from the observation of this one case alone but from the phonetic nature of the sound. Plosives are admittedly the easiest initial sounds to acquire—(i.e. *p, b, t, d, k, g*). The reason is obvious; plosives are emitted with a puff which

[3] Tracy, *Psychology of Childhood,* p. 157, gives the order of learning from first to last, as *p, b, m, h, w, k, d, q, n, t, f, j, s, ch, g, y, sh, v, th, l, r.*

requires relatively less control of facial and laryngeal muscles than other sounds. Next to plosives in this regard stands the lateral (i.e. the *l*). Fricatives and affricates (i.e. all others) require a more delicate control of these muscles. Not only SJ but also her two closest playmates had mastered the lateral *l* long before they had mastered the fricatives *θ* (th) and ð (th).

Of initial vowel sounds, the following were in use at fifteen months:

aɪ	[aɪ]	eye
ɔ	[ˈɔːˈgɔːn]	all gone
o	[ˈoːˈdiː]	orange juice
æ	[ˈæˈpul]	apple
ɛ	[ɛg]	egg
ʌ	[ʌp]	up

This relatively large number of vowel sounds, however, is misleading in that only eight words out of the total 75-word vocabulary contained initial vowels. Five began with the lateral (i.e. *l*) whereas every one of the remaining possessed initial plosives.

The *final* consonants deserve comment. They may be divided into three groups: those weakly stressed, those omitted, and those replaced by vowel sounds. In short, *no final consonants were firmly stressed at this age.* The final consonants present but weakly stressed included only *plosives* (as cu*p*," "boo*k*," "ha*t*") and *nasals* (as "morning" [ˈmɔːnɪŋ] and, "gone" [gɔːn]). To this rule there was but one lone exception—namely of one ending with the lateral consonant (i.e. *l*). This was the word "apple" [ˈæˈpul]. The proper pronunciation of this word was taught SJ only after considerable effort as it was a word in frequent use. The effort spent in mastering it serves to emphasize the natural tendency to articulate only plosives and nasals in final consonant sounds.

Not a fricative or affricative consonant appeared among these final consonant sounds. These sounds whenever they appeared among the words used were either dropped entirely (as [maU] for "mou*th*") or were replaced by vowel sounds (as ['wa:'dI:] for "wat*er*").

At Eighteen Months

The period between the fifteenth and eighteenth months was that of the second "plateau." It was marked by an active increase in the facility of handling old words, but with only a meager increase of new ones. To the vocabulary was added but 21 words; one new initial consonant only appeared — namely v([vɔt] for "what"); not a single new initial vowel appeared; and no new final consonants appeared, although there was a distinct increase in stress upon the old plosive and nasal finals. This led in some cases to the addition of an extra vowel syllable at word endings to aid in this stress—as *hato* ['hæto] for "hat" and *pano* ['þæno] for "pan."

At Twenty-four Months

This period, from the 18th to the 24th months, was one of the outstanding vocabulary growth—marked by a gain of 536 words. This gain was attended by a corresponding increase in the mastery of phonetic sounds.

The following *initial consonants* were acquired:

s
f
j (phonetic dʒ)
sh (phonetic ʃ)
wh (phonetic ʍ)
y (phonetic j)
ch (phonetic tʃ)

These initial consonants, however, were not always used in rapid speech. The *wh*, for example, was often turned into an *f*—as [fi:l] for *wheel*. Strangely enough this *f* was often in turn substituted by a *p*—as [ˈpɪŋgə] for *finger*, and [pɪ*f*] for fish. The reason is obvious. The labia-dental *f* demands a slightly finer muscular activity than the bi-labial *p*.

The following initial consonants had *not* appeared at this age:

r—either dropped or substituted by *w*, as [ˈwaɪˈdɪŋ] for *riding*.

ð (th)—substituted by *d*, as [dæə] for *there* and [əˈnʌdə] for *another*.

θ (th)—substituted by *s*, as [ˈnɪsvɒ] for *nothing*.

z—not found in any vocabulary word; probably could have been articulated; was articulated in final consonants.

The troublesome unmastered phonetic sounds at this age were the *double consonants*. Almost none could yet be mastered. Efforts to pronounce them led either to dropping one of the consonant sounds or to substituting another easier sound in its place. The following are the more prominent instances:

r combinations—*r* usually dropped, as [hu:ə] for *hurry*, [ˈkækə] for *cracker*, [dɪnk] for *drink*.

—substituted in a few combinations by *w*, as [dwɛs] for *dress*, [dwaɪ] for *dry*, and [twi:] for *tree*.

s combinations—s dropped, save in two conditions mentioned below. Typical instances of dropping are [pu:n] for *spoon*, and [ta] for *star*.

—before the lateral and the nasals (i.e. *l* and *m,n*) *s* was supplanted by an aspirate *h*—as [hˈlip] for *sleep*, [hˈlɪpa] for *slipper*, [hˈmo:k] for *smoke*, and [hˈneɪk] for *snake*.

l combinations—*l* dropped in most compounds, as ['a:'peɪn] for *airplane,* [faɪ] for *fly,* ['bɔ'ti:] for *bottle,* and [hɛp] for *help.*
—one exception was *sl* (mentioned under *s*), where the *s* was substituted by an aspirate *h* and the *l* was retained.

All of the common *initial vowels*—namely i, ɪ, ɛ, æ, *a*, ɔ, o, u, ʊ, ʌ, and ə had appeared. The preponderance of consonants in initial sounds, however, was still maintained. By count, only 63 words, or exactly 10 per cent, had initial vowels; the remaining 90 per cent had consonant initiates. There was also a tendency to drop a few initial vowel sounds and start with the first consonant in the word—as [b*a*ut] for *about,* and [gɛnst] for *against.* The reversal of the same principle was also true. That is, an extra consonant often would be inserted before initial vowels of certain words—as [mə'gɛ.n] for *again.* Initial vowels are more difficult than plosive or nasal consonants, since they require an open mouth for initiating a sound. Consequently one finds in the language learning process of children a relatively small percentage of initial vowel sounds.

At Thirty-six Months

The development of ability to master phonetic sounds between the second and third birthdays may be summarized as follows: At thirty-three months the initial consonants, *r* and *z* were properly articulated. At thirty-four months, the last remaining initial consonants, ð (th) and *θ* (th), could be correctly pronounced *when asked for,* but were not correctly sounded in ordinary speech. At thirty-four and one-half months words as *this* and *there* were pronounced correctly but from habit

the word *the* persisted as *de* [di:]. At thirty-six months all *th* sounds were pronounced correctly, except in hasty speech wherein older substitutions for these sounds appeared in a mild form. At thirty-six months also the habit of dropping initial vowel sounds in words, or of substituting them with consonant sounds, had disappeared save in exceptional cases.

The double consonants at this age were uniformly modified toward the correct sound, and all but a few of the *r*, *s*, and *l* combinations had reached the level of correct pronunciation.

SUMMARY

(1) Two "plateau" periods appeared in the vocabulary growth, each one followed by a sudden and rapid increase in the acquisition of new words.

(2) In addition to word learning from objects themselves, a most fruitful source of new words between the ages of one and two years was from children's rhymes, in which the word concept came after the acquisition of the word sound.

(3) A comparison of the parts of speech shows that as the percentage of nouns declined the percentage of verbs increased, indicating the growth of ability to express ideas in complete sentences requiring verbs, rather than in single words or phrases.

(4) The initial consonant sounds appeared in the following order: p, b, m, h, d, w, k, n, t, g, l, s, f, dʒ, ʃ, ʍ, j, tʃ, z, r, ð (th) and θ (th). All but the last four had been acquired at twenty-four months. All had been acquired at thirty-six months.

(5) All of the initial vowel sounds had appeared at twenty-four months, although only 10 per cent of

the words had initial vowel sounds and there was a disposition to substitute consonant sounds for vowels in some words.

(6) At twenty-four months double consonants were still difficult to use, resulting either in dropping of one consonant or in the substituting of one consonant by an easier sound. At thirty-six months most but not all of them had been acquired.

NEWS AND NOTES

Correction—Normal Powell is Assistant Psychologist at Sing Sing Prison and not Psychologist, as stated on page 629 of the October issue of the Journal. James Hargan has held the post of Psychologist since 1930.

The Forty-fourth Annual Meeting of the Ohio Academy of Science will be held at Ohio State University, Columbus, on March 30-31, 1934. Dr. Richard S. Uhrbrock, Director of the Industrial Relations Division of The Proctor & Gamble Company, Ivorydale, Ohio, is the Vice-President of the Section on Psychology.

The Assistant Commissioner for Vocational Education, Dr. J. C. Wright, in the United States Office of Education points out in his annual report that as business resumes, the problem of recovery both for industries and for workers will not be a problem of getting the millions of unemployed workers back on their old jobs or on any other jobs for which they had formerly been fitted, but rather a retraining of workers for new jobs that have become available. This retraining of workers is a prime responsibility of the vocational education system, he feels.

A report of Secretary of the Interior, Harold L. Ickes, shows that more disabled men and women were made self-supporting last year under the Federal Vocational Rehabilitation Act than in any earlier year. Nearly 11,000 disabled men and women have been qualified for employment in some occupation for which their particular physical disability is no longer a handicap. Some 25,300 other physically disabled are now in process of being rehabilitated. Through the cooperation of the National and State vocational rehabilitation services and the National and State Emergency Relief Administration it is expected that a larger proportion of disabled persons will be taken off the public relief rolls and placed in self-supporting employment.

The Detroit Board of Education has recently changed the name of the group of colleges formerly known as the College of the City of Detroit to Wayne University.

The University of Minnesota announces the publication of a bulletin, "Commercial Correspondence Courses and Occupational

Adjustments of Men," which is based upon an analysis of men registered in the Minnesota Employment Stabilization Research Institute. The facts and interpretations regarding cost of courses, mortality rates, and practical values will be of interest to those considering enrollment in commercial correspondence courses. The price of the bulletin is 50 cents.

The following news items have been received from Stanford University: Dr. Bronson Price is now associated with the Psychology Department of the Medico-Biological Institute at Moscow, U. S. S. R. Dr. J. A. McGeoch, of the University of Missouri, is to be Acting Professor at Stanford during the summer of 1934.

BOOK REVIEWS

EDWARD K. STRONG, JR. (with collaborators). *Vocational Aptitudes of Second Generation Japanese in the United States.* Stanford University Publications, University Series, Education-Psychology. v. 1, no. 1, 1933; 183 p.

EDWARD K. STRONG, JR. *Japanese in California.* Stanford University Publications, University Series, Education-Psychology. v. 1, no. 2, 1933; 188 p.

These two monographs are part of a series of four works dealing with the general problem of the Japanese in California. The third number in the monograph series, written by Reginald Bell, is on the *Educational Status of Japanese-Americans in California.* An independent book by E. K. Strong, *The Second Generation Japanese Problem,* will complete the study and will present a summary of the monograph studies together with a discussion of the problem of adjustment of the second generation Japanese in California. These reports of a many sided study of the Japanese, planned and carried out under the direction of Dr. Strong, will give necessary basic sociological, political, economic and psychological data for attempts at the resolution of an important problem of the western states.

The first volume under review is really misnamed. The title reflects Dr. Strong's professional interest more than it does the contents of the book. This is indicated by the table of contents which shows chapters on physical and motor measurements, intelligence, art ability, vocational interests, language usage, personality and character measurements, and delinquency and crime statistics. All of these are based on original data except that on intelligence which is a survey of the literature. As far as one can summarize data in such divergent fields the results indicate Japanese equality with, and in some cases superiority over, the whites, and a definitely greater Japanese superiority over popular opinion of Japanese standing in the various fields.

Because of the large number of comparisons and the small number showing very significant differences it is not necessary to review all of them here. However, the measuring tools might be mentioned because of the interest they may have to other workers not concerned with the primary subject matter. The various abili-

ties were measured thus: Motor skills, Seashore Motor Skills Battery and Miles Reaction Board Tests; art ability, Lewerenz Tests in Fundamental Abilities of Visual Art; vocational interests, Strong's Vocational Interest Blank; personality, Bernreuter Personality Inventory; language ability, honesty and trustworthiness were all studied by rating techniques.

"The general conclusion from the first study is that differences between Japanese and whites, both born in the United States and educated in the same schools of California, are very slight as far as physical (ability), mental and moral traits can be measured. In fact they are so slight that there is no warrant for supposing that Japanese and whites should enter different occupations because of any differences in their physical or mental capacities."

In the second volume of the series are presented for the first time basic statistical data on the social position of the Japanese in California. These data are based on personal interviews with some 9,000 individuals, men, women and children, representing 9.7 per cent of the total Japanese population of the state. Additional evidence is presented from public reports, previous investigations and other documentary sources. The sociological data include: distribution of the Japanese population, vital statistics, education, occupation, income and value of property, and religion. It is found that, while about one-half of all Japanese are second generation, only 6 per cent are of voting age; about one-half are engaged in agriculture; the average number of children is slightly more than 3 per family; the education of those who are not in school compares favorably with the whites, averaging around 7-8 grades. These suggestive selections with many more data support the author's statement:

"The evidence in this study is that the first generation have progressed satisfactorily. In fact the complaint has been that they were altogether too successful. The evidence, furthermore, is that the second generation are progressing satisfactorily in school work and that they are planning upon careers in about the same manner as white youths."

There is no doubt that these studies are important contributions not only to a local race problem, but also to scientific study of the psychology and sociology of race differences. Dr. Strong is to be commended on having brought to a successful conclusion a task which, observed only in the final product, is so stupendous as to be appalling.

C. M. Louttit,
Indiana University.

CABOT, RICHARD C. *The Meaning of Right and Wrong.* The Macmillan Company, New York, 1933. X + 463 pp.

The writer finds the foundations of ethics in the keeping of agreements, which may be tacit or explicit. Illustrating by Conrad's story, Lord Jim in deserting his ship failed to keep an agreement which was tacitly made by the acceptance of the life of a seaman. Agreements are authenticated by desires and are to be judged in terms of whether they express the genuine desires of those concerned in the agreement. When they fail to do so, as when they are impulsively made to express impulsive desires, or when they become obsolete in the light of changed conditions, they may be replaced by more adequate agreements with the consent of those concerned. The need for growth is the authoritative need and the one in terms of which agreements have their greatest validity.

The validity of customary and legal agreements, the interpretation of agreements, and the meaning and limits of due care are among the applications discussed.

"Wrong" is the subject of one division of the book, and here we find self deceit discussed at length, and recklessness as a form of self deceit.

Supermorality is "action done for the love of some one or something, that is, from an ardent sense of need in us or for us." It is to be found in heroism and creative activity. It is called out by our strength, while morality is called out by our weakness.

In appendices the author touches upon some unsolved problems of ethics, including the problem of eugenics. Here he also gives a short list of definitions of terms, and analyzes briefly the fifty-four declarations of war made between 1914 and 1918.

The work is original in conception, and is clearly and interestingly presented. Principles are greatly clarified by illustrative material drawn largely from literary and legal sources.

AMOS C. ANDERSON,
Ohio University.

GATES, A. I., and BENNETT, C. C. *Reversal Tendencies in Reading.* New York, Teachers College, 1933, pp. 33.

The tendency toward reversing the sequence of letters within words has long been recognized as an important diagnostic sign of reading deficiency. This monograph is a contribution on the theoretical and practical significance of such reversal errors. Previous work is criticised, new experimental findings presented, and remedial suggestions offered. A comparison of children showing marked tendencies toward reversals with those revealing minimum

tendencies seems to indicate that left-eyedness, left-handedness, left-eyedness combined with right-handedness, and visual defects are associated in some degree with reversal tendencies. The authors tend to favor an "acquired habit" theory rather than a "brain organization" theory to explain the reversals.

As a whole the monograph is a very worth while contribution to diagnostic and remedial reading. There are, however, certain statements and inferences whose adequacy may be questioned. In considering Hinshelwood's theory of reading disability, the author's statement that "the prognosis of such cases, logically, would not be hopeful" is misleading. Although Hinshelwood's theory may be unsatisfactory, he himself states (p. 90) that "in nearly all cases of congenital word-blindness . . . the children . . . can be taught to read." Furthermore, his remedial treatment was a worthy forerunner to the present successful methods.

Throughout the report repeated emphasis is placed upon the development of orderly sequences of eye movements from left to right along words and lines of print. It is suggested that the finger be moved along the line, or a moving spot of light be projected along the line by a motion picture machine to train the eye movements. (This view is just opposite that taken by the senior author in "The Improvement of Reading," p. 234.) These highly formal techniques for training eye movements are unnecessary and may in many instances become a liability. Any training to promote orderly sequences of perception should be made an intrinsic part of the reading situation as is done in many of the authors' excellent suggestions. Then the eye movements as such will take care of themselves.

A careful consideration of the part played by eye movements in reading would suggest that the authors have made questionable inferences concerning the part played by the eye movements in enhancing tendencies to reversals. It is well known that all beginners in reading show tendencies toward reversals. Satisfactory evidence in terms of eye-movement habits is not adduced by the authors to explain why these reversals persist in only the few children who become deficient in reading.

MILES A. TINKER,
University of Minnesota.

GARRETT, HENRY E. and SCHNECK, MATTHEW R. *Psychological Tests, Methods, and Results.* New York, Harpers, 1933, pp. x + 137 + 235.

As a vast amount of research is constantly being attempted in the field of psychological testing and many workers are devoting

their entire time to the development of these tests, this combination textbook and handbook will be a welcome addition to the libraries of students and clinicians. In this one volume, which is really two books, the authors have succeeded in discussing a wide variety of tests briefly, objectively, and concisely.

Part One, *The Measurement of Simpler Functions*, is largely the work of Schneck, and Part Two, *The Measurement of Complex Functions*, was written entirely by Garrett. These writers have attempted to give the student a perspective of all types of mental tests, allotting space to each, according to its relative importance. The general arrangement of the book is such that the information regarding each test is easy to find and may be studied without reading unrelated material.

The first part consists of four chapters dealing respectively with tests of physical and sensory capacity, motor ability and mechanical aptitude, perception and attention, and learning, association and memory. In all, some fifty-odd tests are discussed and a bibliography of 252 titles is distributed at the end of the chapters. In discussing each test, the author gives a description of apparatus, methods of administering, keeping the record, norms, and from whom the apparatus or tests are available. He, also, cites certain investigations in which the tests have been used and gives the results obtained.

The general plan of Part II is the same as that of Part I, but here we find more discussion added to the description and explanation of the tests and methods. In this section there are five chapters in which the author takes up verbal tests of "general intelligence," performance and non-language tests, the measurement of personality and temperament, tests in special fields, and applications of psychological tests. Here a generous list of references is also found at the end of each chapter.

In the 65 pages of the first chapter (Part II), the author discusses most individual and many group tests of "general intelligence" and devotes a few pages to tests of educational achievement. He says that the so-called "intelligence test" is really a test of scholastic ability and for purposes of educational prognosis should be accompanied by tests designed to measure achievement and ability in school subjects. His criticism of the use of mental age and IQ ratings in group tests of intelligence explains very clearly the serious faults to be found with this method and should be convincing even to its most ardent users.

After examining the results obtained with performance and non-language tests, the conclusion is reached that these should not be substituted for the verbal type except with young children

and low grade adults. The author fully appreciates the shortcomings of personality measurement and has succeeded very well in giving the available facts regarding the progress made so far without being either gullible or hyper-critical.

Under tests in special fields are included tests of music, art, commercial, vocational, and professional aptitude, and in the final chapter he discusses the many and varied uses which psychological tests have found in the study of individual and group differences.

In any book of this kind a few typographical errors and unwarranted assumptions are bound to creep in, but in the present work we find only one or two such instances which will, no doubt, be corrected in subsequent editions.

The feeling which one has after completing this excellent survey of the entire field of testing is, that while clearly showing what has already been done, it will help to point out the way for further advancement. Its major contribution, however, will be to the student who wishes to gain a generalized background in testing before taking up the use of any particular kind of tests.

H. C. MAHAN,
Warren (Pa.) *State Hospital.*

NEW BOOKS AND PAMPHLETS

Books and pamphlets for review should be sent to James P. Porter, Editor, JOURNAL OF APPLIED PSYCHOLOGY, Ohio University, Athens, Ohio.

Applied Psychology. RICHARD WELLINGTON HUSBAND. Harper & Brother, New York City, 1934. $2.90. 654 pp.

Broadcasting in the United States. National Association of Broadcasters, Philip G. Loucks, Managing Director, Washington D.C., 1933. 191 pp.

Die Charaktereigenschaften. FRANZISKA BAUMGARTEN. Verlag A. Francke, Bern. 81 pp.

Handbook of Child Psychology. Second Edition Revised. Edited by Carl Murchison, Clark University Press, Worcester, Mass. 956 pp.

Intelligence and Scholastic Attainment. H. T. PARKER. Melbourne University Press, Melbourne, Australia, 1932. 3 sh. 6 pence. 64 pp.

Interpreting the Secondary School of the Public. BELMONT FARLEY. Bulletin No. 17, 1932, U.S. Office of Education, Washington, D.C. 10 cents. 113 pp.

An Introduction to Animal Psychology. NORMAN L. MUNN. Houghton Mifflin Company, New York City, 1933. 439 pp.

Japanese in California. EDWARD K. STRONG. Stanford University Press, Stanford University, Calif. $1.00. 188 pp.

Keeping a Sound Mind. JOHN J. B. MORGAN. Macmillan Company, New York, 1934. 440 pp.

Knowing and Helping People. HORATIO W. DRESSER. The Beacon Press, Boston, 1933. $2.50. 268 pp.

Part-Time Secondary Schools. GRAYSON N. KEFAUVER, VICTOR H. NOLL, and C. ELWOOD DRAKE. Bulletin No. 17, 1932. U.S. Office of Education, Washington, D.C. 10 cents. 98 pp.

Prediction of Vocational Success. EDWARD L. THORNDIKE. The Commonwealth Fund, New York. $2.50. 284 pp.

Report of the President of Columbia University. Columbia University, New York. 61 pp.

Textbook of Normal Psychology. ROY M. DORCUS and G. WILSON SHAFFER. Williams & Wilkins Company, Baltimore, Md., 1934. $4.00. 389 pp.

Twenty-eighth Annual Report of the President and Treasurer of the Carnegie Foundation for the Advancement of Teaching. June 30, 1933.

Vocational Aptitudes of Second Generation Japanese in the United States. EDWARD K. STRONG, REGINALD BELL and others. Stanford University Press, Stanford University, Calif. $1.00. 183 pp.

Directory of American Psychological Periodicals

(Continued)

Psychological Abstracts—Eno Hall, Princeton, N. J. American Psychological Association. Subscription $6.00. 700 pages ann. Edited by W. S. Hunter. Monthly. Abstracts of psychological literature. Founded 1927.

Pedagogical Seminary and Journal of Genetic Psychology—Worcester, Mass.: Clark University Press. Subscription $14.00 per annum; $7.00 per volume. 1,000 pages annually (2 volumes). Edited by Carl Murchison and an international cooperating board. Quarterly. Child behavior, animal behavior, and comparative psychology. Founded 1891.

Journal of Social Psychology—Worcester, Mass.: Clark University Press. Subscription $7.00. 500-600 pages annually. Edited by John Dewey and Carl Murchison. Quarterly. Political, racial, and differential psychology. Founded 1929.

Personnel Journal—Baltimore: The Williams & Wilkins Co. Subscription $5.00. 500 pages, approximately. Edited by Walter V. Bingham. Bi-Monthly.

Mental Measurement Monographs—Baltimore: The Williams & Wilkins Company. Subscription $5. 400 pages annually. Edited by Knight Dunlap and Buford Johnson. Published without fixed dates, each number a single research. Founded 1925.

Studies in Psychology and Psychiatry—Baltimore: The Williams & Wilkins Company. Subscription $5.00. 400 pages annually. Edited by Edward A. Pace. Published without fixed dates, each number a single research. Founded 1926.

Journal of Philosophy—New York: 515 West 116th Street. Subscription $4.00. 728 pages per volume. Founded 1904. Bi-weekly. Edited by F. J. E. Woodbridge, Wendell T. Bush, and H. W. Schneider.

Child Development—Baltimore, Md.: Williams & Wilkins Company. Subscription $5.00. 400 pages annually. Edited by Buford Johnson. Quarterly. Founded 1930.

Character and Personality—Durham, N. C.: Duke University Press. Subscription $2.00 per annum. Quarterly. Edited by Robert Saudek (London) and Ernest Seeman (Durham, N. C.). Founded 1932

Directory of American Psychological Periodicals

American Journal of Psychology—Ithaca, N. Y.: Cornell University. Subscription $6.50. 624 pages annually. Edited by M. F. Washburn, K. M. Dallenbach, M. Bentley, E. G. Boring. Quarterly. General and experimental psychology. Founded 1887.

Journal of General Psychology—Worcester, Mass.: Clark University Press. Subscription $14.00 per annum; $7.00 per volume. 1,000 pages annually (2 volumes). Edited by Carl Murchison and an international cooperating board. Quarterly. Experimental, theoretical, clinical, and historical psychology. Founded 1927.

Psychological Review—Princeton, N. J.: Psychological Review Company. Subscription $5.50. 540 pages annually. Bi-monthly. General Psychology. Edited by Howard C. Warren. Founded 1894.

Psychological Monograph—Princeton, N. J.: Psychological Review Company. Subscription $6.00 per vol. 500 pp. Ed. by Raymond Dodge. Published without fixed dates, each issue one or more researches. Founded 1895.

Psychological Index—Princeton, N. J.: Psychological Review Company. Subscription $4.00. 300-400 pp. Edited by W. S. Hunter. An annual bibliography of psychological literature. Founded 1895.

Psychological Bulletin—Princeton, N. J.: Psychological Review Company. Subscription $6.00. 720 pages annually. Psychological literature. Monthly. Edited by Edward S. Robinson. Founded 1904.

Archives of Psychology—New York: Columbia University. Subscription $6. 500 pp. ann. Ed. by R. S. Woodworth. Published without fixed dates, each number a single experimental study. Founded 1906.

Journal of Abnormal and Social Psychology—Eno Hall, Princeton, N. J. American Psychological Association. Subscription $5. Foreign, $5.25. Editor, Henry T. Moore. Quarterly. 448 pages annually. Abnormal and Social. Founded 1906.

Psychological Clinic—Philadelphia: Psychological Clinic Press. Subscription $3.00. 288 pages. Ed. by Lightner Witmer. Without fixed dates (9 numbers). Orthogenics, psychology, hygiene. Founded 1907.

Psychoanalytic Review—Washington, D. C.: 3617 10th St., N. W. Subscription $6. 500 pages annually. Psychoanalysis. Quarterly. Ed. by W. A. White and S. E. Jelliffe. Founded 1913.

Journal of Experimental Psychology—Princeton, N. J. Psychological Review Company. 500 pages annually. Experimental. Subscription $6.00. Bi-monthly. Edited by S. W. Fernberger. Founded 1916.

Journal of Applied Psychology—Indianapolis: The C. E. Pauley & Co., Inc. Subscription $5.00. 600 pages annually. Bi-monthly. Edited by James P. Porter, Ohio University, Athens, Ohio. Founded 1917.

Journal of Comparative Psychology—Baltimore: The Williams & Wilkins Co. Subscription $5.00 per volume of 450 pages. Three volumes every two years. Edited by Knight Dunlap and Robert M. Yerkes. Founded 1921.

Comparative Psychology Monographs—Baltimore: Johns Hopkins University Press. Subscription $5. 400 pages per volume. Knight Dunlap, Managing Editor. Published without fixed dates, each number a single research.

Genetic Psychology Monographs—Worcester, Mass.: Clark University Press. Subscription $7.00 per volume. Two volumes per year, 600 pages each. Edited by Carl Murchison. Monthly. Each number one complete research. Child behavior, animal behavior and comparative psychology. Founded 1925.

(Continued on Third Cover)